ISSN: 2150-4091 Volume 2, Number 6, June 2010

Scientific Research

# Natural Science

ISSN: 2150-4091



www.scirp.org/journal/ns/

# Journal Editorial Board

ISSN: 2150-4091 (Print) ISSN: 2150-4105 (Online)
http://www.scirp.org/journal/ns/

Vol.2, No.6, 527-666 (2010) Natural Science

# TABLE OF CONTENTS

## Volume 2, Number 6, June 2010

# Natural Science

# Journal Information

## SUBSCRIPTIONS

The *Natural Science* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

### Subscription rates:

Print: $50 per copy.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

## SERVICES

**Advertisements**
Advertisement Sales Department, E-mail: service@scirp.org

**Reprints (minimum quantity 100 copies)**
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

## COPYRIGHT



## PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: ns@scirp.org

**Vol.2, No.6, 527-534 (2010)**
doi:10.4236/ns.2010.26066

**Natural Science**

# Differentiation of wild boar and domestic pig populations based on the frequency of chromosomes carrying endogenous retroviruses

**Sergey V. Nikitin[1], Nikolay S. Yudin[1], Sergey P. Knyazev[2*], Ruslan B. Aitnazarov[1], Vitaliy A. Bekenev[3], Valentina S. Deeva[3], Galina M. Goncharenko[3], Victor F. Kobzev[1], Margarita A. Savina[1], Viktor I. Ermolaev[1]**

[1]Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
[2]Novosibirsk State Agrarian University, Novosibirsk, Russia; *Corresponding Author: knyser@rambler.ru
[3]Siberian Research Institute of Animal Husbandry Siberian Branch, Russian Academy of Agricultural Sciences, Krasnoobsk, Russia

Received 25 March 2010; revised 30 April 2010; accepted 13 May 2010.

## ABSTRACT

**Analysis of the frequencies of chromosomes carrying various classes of porcine endogenous retroviruses (PERVs) and combinations of these classes was performed in the swine species *Sus scrofa L.* by using maps constructed in two principal component coordinates. Four population clusters can be recognized in the maps. Cluster 1 is formed by wild boars,cluster 2 by domestic meat breeds, cluster 3 by meat-and-lard (universal) breeds, and cluster 4 by miniature pigs. The maps indicate that modern domesticated swine meat breeds are the closest to the wild type. Meat-and-lard domestic swine breeds are more distant from wild boars, and miniature pigs are diverged the most. The maps showed that microevolution processes associated with PERV carriership frequency had two basic dimensions, or vectors: the vector of fat deposition variation and the "minus" selection vector (determination of commercial traits). Thus, PERVs may cause variation in pig physiology.**

**Keywords:** *Sus scrofa*; Endogenous Retroviruses; Pig Genetics; Microevolution; Genetic Distance Determination

## 1. INTRODUCTION

Porcine endogenous retroviruses (PERVs) became an integral part of swine genomes, including *Sus scrofa* L. 1758 (*Suidae*, Mammalia), before the formation of the *Sus* genus. This is confirmed by their presence in bush-pigs (*Potamochoerus larvatus* and *P. porcus*) and wart-hogs (*Phacochoerus africanus*) [1]. Three PERV classes are known: A, B, and C. Different classes are highly similar in the nucleotide sequences of the *gag* (group-specific antigens) and *pol* (polymerase) genes but differ in the nucleotide sequence of the receptor-binding domain of the *env* (envelope) gene, which encodes the envelope protein of the virus [2-4]. This difference is responsible for the host range in various virus classes.

Porcine endogenous retrovirus copies carried by different pig varieties are distinct in nucleotide composition, expression, and ability to produce infectious virions. It is believed that the pig genome can carry 6-10 replication-competent proviruses, 30-50 full-size PERV copies, and 100-200 loci carrying truncated virus sequences [1]. Comprehensive studies of pigs belonging to the Large White breed were undertaken to evaluate the number of genomic sequences coding for full-length replication-competent proviruses [5-7]. As a result, significant variations in the distribution and number of proviruses were found in this breed. Viral genomes were also analyzed in the following breeds: Westran [8,9], Duroc, Landrace, Yorkshire, Berkshire, and their hybrids [10,11]; Chinese breeds Banna miniature pig, Wu-Zhi-Shan, Nei Jiang [12], and Meishan [11]; and west European wild boars [9]. In these breeds, as in Large White pigs, PERV sequences are dispersed throughout the genome. The breeds differ in the PERV copy number, chromosomal distribution, and presence of full-length sequences. These traits also varied within the breeds.

Previously, differences in the prevalence of individuals with chromosomes carrying PERVs of various classes and their combinations between domestic pig breeds,

between wild and domestic pigs, and between wild pigs of East Europe and Central Asia were demonstrated [13,14].

In this work, we analyze the differentiation between populations of wild and domestic pigs. For this purpose, we performed statistical assessment of the population frequencies of chromosomes carrying certain PERV classes and combinations is using maps constructed in two principal component coordinates.

## 2. MATERIALS AND METHODS

Experiments were performed with blood samples from three subspecies of wild boars, five commercial breeds of domestic pigs, and one breed of laboratory miniature pigs (**Table 1**). Wild boar animals of the European *Sus scrofa scrofa* variety (SSS) were obtained from the Voronezh Biosphere Reservation. Wild boars of the Romanian subspecies *S. s. attila* were taken from two southern Ukrainian populations (SAS and SAN). The Central Asian Wild Boar subspecies *S. s. nigripes* (SSN) was represented by animals hunted down in the Chu Valley, Kyrgyzstan. Domestic pigs *Sus scrofa domestica* of the Large White breed included animals of the Achinsk (LWA) and Novosibirsk (LWN) types bred at the Inya stud Farm. Landrace pigs were obtained from the Kudryashovskoe farm (LNK) and the Experimental Farm of the Siberian Branch of the Russian Academy of Sciences, hereafter referred to as the Experimental Farm (LNE). Duroc pigs were obtained from the Kudryashovskoe farm (DRK). Animals of the SM1 precocious meat breed were obtained from the Tulinskoe work-study unit. The Kemerovo breed included animals from the Yurginskii breeding farm (KMR). Miniature pigs (MS) were obtained from the Experimental Farm. A total of 636 blood samples were studied: 35 from mature wild boar males and females and 601 from domestic pigs (mature breeding males and females and youngsters below two months). Blood was taken from the anterior vena cava of living domestic pigs or from the heart of killed wild boars.

DNA was isolated from blood samples and analyzed for the presence of PERV classes by polymerase chain reaction (PCR) [11]. The sequences of primers complementary to the *env* gene sequences of various PERV classes were adopted from the literature. Primers for *env*A: forward 5′-TGG AAA GAT TGG CAA CAG CG-3′ and reverse 5′-AGT GAT GTT AGG CTC AGT GG-3′ [3]. Primers for *env*B: forward 5′-TTC TCC TTT GTC AAT TCC GG-3′ and reverse 5′-TAC TTT ATC GGG TCC CAC TG-3′ (ibid.) Primers for *env*C: forward 5′-CTG ACC TGG ATT AGA ACT GG-3′ and reverse 5′-ATG TTA GAG GAT GGT CCT GG-3′ [4].

The population frequencies of chromosomes carrying various PERV classes and their combinations were determined by Bernstein method modified for a gene with multiple copies located on different chromosomes [14].

Phylogenetic relationships among the populations under study were studied in maps constructed in two principal component coordinates. Determination of genetic distances was followed by sample ordination [15]. After scaling, each population was defined as a point in a 500 × 500 arbitrary unit area. Three types of genetic distances were used: 1) Euclidean distances $D_1 = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$ ,

**Table 1.** Populations of wild boars and domestic pigs.

| Subspecies or breed, locality | Designation | Sample size |
|---|---|---|
| Wild boar *S. s. scrofa*, Voronezh Biosphere Reservation | SSS | 12 |
| Wild boar *S. s. attila*, Carpathians (Ukraine) | SAC | 12 |
| Wild boar *S. s. attila*, Ukraine, Nikolaev Region | SAN | 7 |
| Wild boar *S. s. nigripes*, Kyrgyzstan | SSN | 4 |
| Large White breed, Novosibirsk type, Inya farm | LWN | 101 |
| Large White breed, Achinsk type, Inya farm | LWA | 99 |
| Kemerovo breed, Yurginskii breeding farm | KMR | 165 |
| Landrace, Kudryashovskoe farm | LNK | 15 |
| Landrace, Experimental Farm | LNE | 30 |
| SM1 precocious meat breed, Tulinskoe work-study unit | SM1 | 21 |
| Duroc, Kudryashovskoe farm | DRK | 10 |
| Miniature pigs, Experimental Farm | MS | 160 |

where n is the number of phenotypes, and $p_i$ and $q_i$ are frequencies of phenotypes in populations to be compared [16]; 2)Harpending–Jenkins distances $D_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(p_i - \dot{p})(q_i - \dot{p})}{\dot{p}(1-\dot{p})}$, where $\dot{p}$ is the weighted mean frequency of a certain phenotype in populations to be compared, and n is the number of genes according to which the populations are compared [17]; 3) Nei's distances $D_3 = -\ln J_{pq} + 1/2(\ln J_p + \ln J_q)$, where $J_p = \sum_{i=1}^{n} p_i^2$, $J_q = \sum_{i=1}^{n} q_i^2$, $J_{pq} = \sum_{i=1}^{n} p_i^2 q_i^2$, $J_p$ is the theoretical homozygosity in the first population, $J_q$ is the theoretical homozygosity in the second population, and $J_{pq}$ is the mutual identity of the populations under comparison [16].

Two models were considered for construction of maps in two principal component coordinates. In model M-1, the frequencies of chromosomes carrying PERV classes were presented as frequencies of three independent factors with two variables: *env*A$^+$ and *env*A$^-$, *env*B$^+$ and *env*B$^-$, *env*C$^+$ and *env*C$^-$. In model M-2, the frequencies of chromosomes carrying PERV class combinations were presented as frequencies of seven independent factors with two variables: *env*A$^+$ and *env*A$^-$, *env*B$^+$ and *env*B$^-$, *env*C$^+$ and *env*C$^-$, *env*AB$^+$ and *env*AB$^-$, *env*AC$^+$ and *env* AC$^-$, *env*BC$^+$ and *env*BC$^-$, *env*ABC$^+$ and *env*ABC$^-$. For each model, three versions of maps of the phylogenetic relationship among the varieties were constructed.

## 3. RESULTS

As shown in our previous study [14], frequencies of chromosomes carrying various PERV classes and their combinations vary significantly among the subspecies of wild boars and breeds of domesticated pigs (*Sus scrofa* L. 1758) (SUIDAE, MAMMALIA), as well as among herds within a breed. One to three distinct PERV classes were detected in chromosomes of the populations under study. In addition to single provirus copies, such chromosomes contained combinations AB, AC, BC, and ABC (**Table 2**).

Maps in two principal component coordinates constructed on the base of models M-1 and M-2 demonstrate features of the phylogenetic relationship between populations determined by microevolutionary processes (**Figure 1**). It should be emphasized that model M-1, which considers the frequencies of chromosomes carrying certain PERV classes, and model M-2 that deals with the frequencies of chromosomes carrying combinations of these classes, yield different results (**Figure 1**).

Four population clusters can be recognized in maps constructed on the base of frequencies of chromosomes carrying PERV classes. Cluster 1 is formed by wild boars; cluster 2 by domestic meat breeds; cluster 3 by meat-and-lard (universal) breeds, and cluster 4 by miniature pigs (**Figure 1**). These clusters form a certain logical order. Three clusters form one straight line: wild boars, meat breeds, and universal breeds. The fourth cluster, miniature pigs, is distant from this line. Thus, according to the M-1 model, the frequencies of PERV classes show the following trend associated with morphotypes: wild boar → commercial meat morphotype → commercial meat-and-lard morphotype. The maps also indicate that variation among populations within the

**Table 2.** Frequencies of chromosomes carrying certain PERV classes and combinations of these classes in wild boar and domestic pig populations. The classes are identified by *env* gene sequences.

| Population* | Frequencies of PERV class carriers | | | Frequencies of PERV class combination carriers | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *env*A | *env*B | *env*C | A | B | C | AB | AC | BC | ABC |
| SSS | 0.011 | 0.023 | 0.008 | 0.000 | 0.012 | 0.000 | 0.003 | 0.000 | 0.000 | 0.008 |
| SAC | 0.000 | 0.039 | 0.027 | 0.000 | 0.012 | 0.000 | 0.000 | 0.000 | 0.027 | 0.000 |
| SAN | 0.000 | 0.029 | 0.013 | 0.000 | 0.016 | 0.000 | 0.000 | 0.000 | 0.013 | 0.000 |
| SSN | 0.008 | 0.004 | 0.020 | 0.008 | 0.004 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 |
| LWN | 0.068 | 0.047 | 0.005 | 0.064 | 0.041 | 0.000 | 0.001 | 0.000 | 0.005 | 0.000 |
| LWA | 0.065 | 0.114 | 0.013 | 0.000 | 0.045 | 0.000 | 0.060 | 0.005 | 0.008 | 0.000 |
| KMR | 0.087 | 0.079 | 0.037 | 0.000 | 0.006 | 0.000 | 0.055 | 0.019 | 0.005 | 0.013 |
| LNK | 0.042 | 0.034 | 0.008 | 0.041 | 0.034 | 0.008 | 0.000 | 0.000 | 0.000 | 0.000 |
| LNE | 0.052 | 0.052 | 0.036 | 0.000 | 0.000 | 0.000 | 0.024 | 0.007 | 0.007 | 0.022 |
| SM1 | 0.028 | 0.051 | 0.051 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.023 | 0.028 |
| DRK | 0.042 | 0.018 | 0.009 | 0.042 | 0.018 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 |
| MS | 0.109 | 0.099 | 0.109 | 0.000 | 0.000 | 0.000 | 0.000 | 0.010 | 0.000 | 0.099 |

*Designations follow **Table 1**.

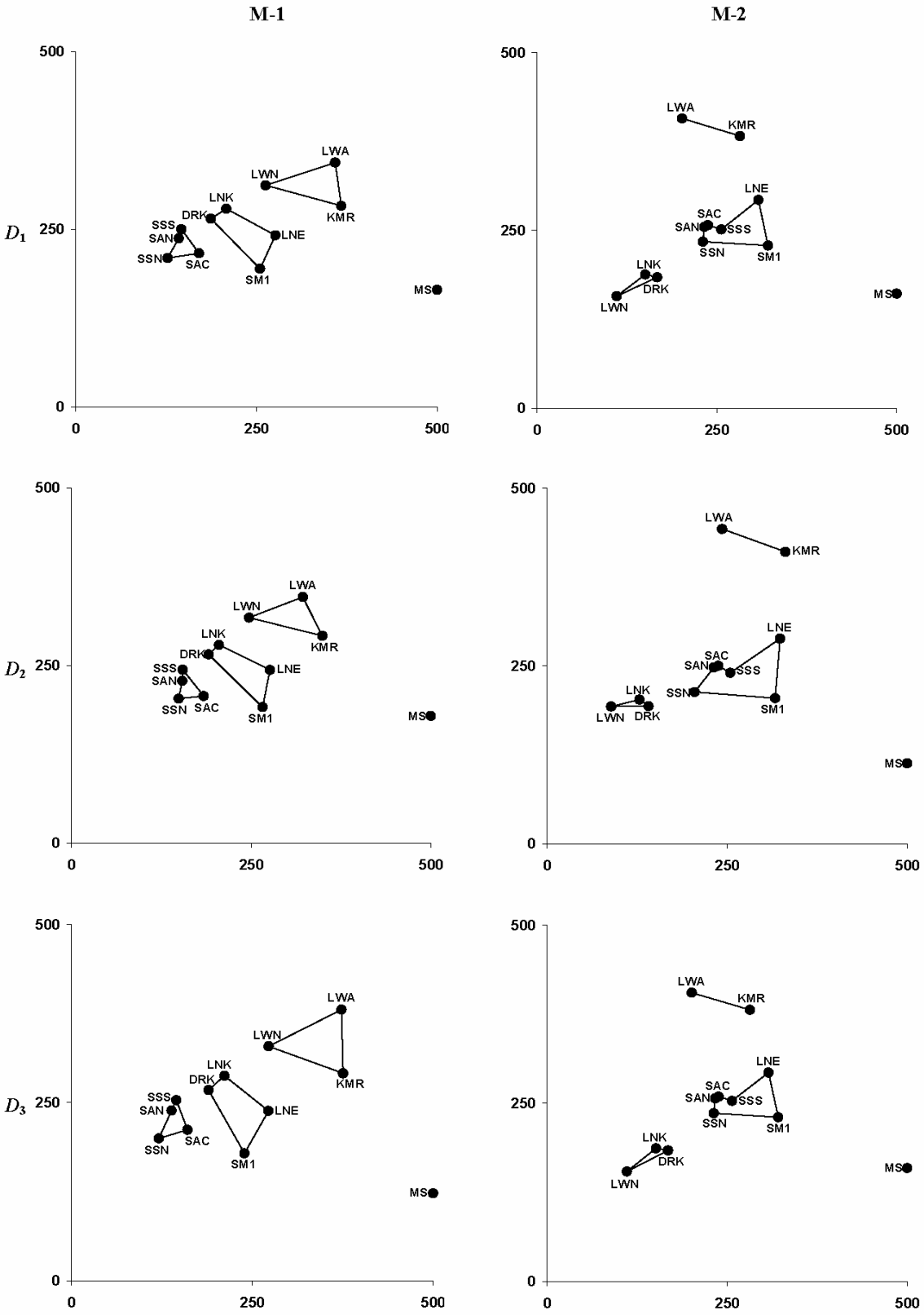



**Figure 1.** Maps constructed in two principal component coordinates on the base of frequencies of chromosomes carrying PERV classes (M-1) and class combinations (M-2). Designations: $D_1$, $D_2$, $D_3$ are genetic distances: Euclidean, Harpending–Jenkins', and Nei's ones, respectively. Population designations follow **Table 1**.

clusters is nonrandom. The vector of this variation is directed to point 0, 500 (bottom-right corner of the map). Thus, it is reasonable to suggest that cluster 4 (miniature pigs) is the farthest deviation from this vector and that these pigs originated from a population initially belonging to cluster 3 (meat-and-lard morphotype).

A markedly different pattern is seen in the maps constructed on the base of frequencies of chromosomes carrying PERV class combinations (**Figure 1**). It is similar to that obtained by using the M-1 model (**Figure 1**) in that the miniature pig population is distant from other populations, whereas wild boar populations still form a compact cluster. However, the clusters of meat and universal meat-and-lard breeds are extended in parallel to the 0, 0; 500, 500 line, so that the populations belonging to these clusters are located on the opposite sides of the wild boar cluster (**Figure 1**). The populations can be combined into clusters according to their locations in the maps (**Figure 1**). Cluster 1 includes Landrace and Duroc domestic pigs from the Kudryashovskoe farm and the Novosibirsk subbreed of Large White pigs. Noteworthy, these pigs have a history of the most intensive selection for meat yield. Cluster 2 includes Large White pigs of the Achinsk subbreed and the Kemerovo breed. These populations belong to the universal meat-and-lard morphotype. Cluster 3 is formed by miniature pigs. Cluster 4, located in the centres of the maps, is of special interest. It includes wild boars, SM-1 pigs, and Landrace pigs from the Experimental Farm. This combination is reasonable. The intensity of selection for meat yield in these two populations was less than in populations of the first cluster, and this is the cause of the reversion to the wild morphotype observed in these populations. It is known that meat breeds had been raised from meat-and-lard and lard morphotypes [18]; therefore, they deviated even more from wild boars. In Landrace pigs of the Experimental Farm, this shift is directed to the cluster of meat-and-lard populations, and in the SM-1 breed to miniature pigs, which can be considered closer to their ancestral Asian lard breeds [19,20]. The Kemerovo breed is located between the Achinsk Large White subbreed and miniature pigs in both maps. The most likely cause of this location is that the Kemerovo breed was at first raised as a lard breed [21], and later selection was directed to the universal meat-and-lard type [22]. Thus, maps constructed on the frequencies of chromosomes carrying PERV class combinations reveal finer features of population differentiation than maps of simple PERV class carriership. These features are associated with the differentiation among populations within large groups, such as wild boars and commercial domestic pig morphotypes, rather than with the differentiation among the groups.

There were two questions to be raised in our work: what wild boar population is closer to the present-day domestic pig according to PERV prevalence and what domestic pig breeds are closer to wild boars? To answer these questions, we employed maps constructed in two principal component coordinates according to two models: a model named M-1 that considers frequencies of PERV classes, and a model M-2 that deals with various PERV class combinations. The distances between populations seen in the maps (**Figure 1**) are presented as bar graphs (**Figure 2**). The graphs obtained on the base of the M-1 model show that the European wild boar *Sus scrofa scrofa* subspecies from the Voronezh Biosphere Reserve is the closest to the domestic pig, and the Central Asian wild boar subspecies *Sus scrofa nigripes* is the farthest. The distances determined on the base of the M-2 model show the same result, although less clearly. Both models indicate that modern domesticated swine meat breeds are the closest to the wild type. Meat-and-lard domestic swine breeds are farther from wild boars, and miniature pigs are the farthest.

## 4. DISCUSSION

Different aspects of PERV-associated features of differentiation of swine (Sus scrofa) populations were revealed by using the models presented in this study. Previous results [13,14] were summarized, and hypotheses concerning the role of PERVs in microevolutionary changes occurring in Sus scrofa populations were substantiated. The variation in the frequencies of chromosomes carrying certain PERV classes and class combinations follows two microevolutionary vectors. The first vector is more specific: wild boars → meat pig breeds → meat-and-lard pig breeds. It can be defined as an increase in fat deposition rate in pig subspecies and breeds. The second vector is of a more general nature: wild boars and commercial domestic pig breeds → miniature pigs. It is a “minus” selection vector. It can be suggested that in the first case PERVs tag loci affecting fat deposition rate and in the second case are associated with loci controlling some commercial and adaptive traits. However, the following hypothesis appears to be more plausible: retroviral copies are inserted into certain functionally important genome regions, thereby disrupting the normal function of genes located in the insertion sites or nearby. These aberrations may give rise to undesirable or, to the contrary, desirable traits. Thus, the vector of increasing fat deposition appears to be a special manifestation of the “minus” selection vector. The latter should be more accurately termed the genomic aberration vector. This hypothesis is in agreement with the formerly reported data that the frequencies of individuals and

chromosomes carrying PERVs in populations naturally selected for fitness (wild boars) or intensively selected for pork production (pigs of the Kudryashovskoe herd) are much lower than in populations undergoing a less intensive selection [14]. Also, it was shown that a class B PERV copy was inserted into the BAT1 gene (coding for RNA helicase) of Large White pigs [9]. It should be mentioned that the frequency of PERV carriers was higher among miniature pigs, the product of "minus" selection. They can be considered a pathological form because of their slow growth, small size, low fertility, high postnatal mortality, and a tendency to obesity. Note that in miniature pigs not only the highest PERV carriership frequency was recorded but also the highest frequency of chromosomes carrying all the three PERV classes in comparison to all other populations investigated (**Table 2**).

The proximity of wild boars from the Voronezh Biosphere Reserve located in the centre of European Russia to present-day domestic pigs (**Figure 2**) appears to be due to convergence rather than divergence. The Voronezh wild boar population was formed by a recent natural crossing between the subspecies *Sus scrofa scrofa* and *Sus scrofa attila* during migrations of wild boars from the West (from Europe) and South (from Ukraine and Caucasus) [23,24]. Therefore, its early assignment to *S. s. scrofa* is entirely formal. Both these subspecies were among the ancestors of modern European breeds of *Sus scrofa domestica* [23,25,26]. Thus, the similarity between these two forms, the wild boar and domestic pig that originate from common ancestral subspecies is natural and of a convergent nature.

It is reasonable to suggest that the graph series obtained by the ranking similarity of domestic pig populations to wild boar are related to the vector of genomic changes induced by PERVs. Breeds of the meat type are in the closest proximity to wild boars. They were raised

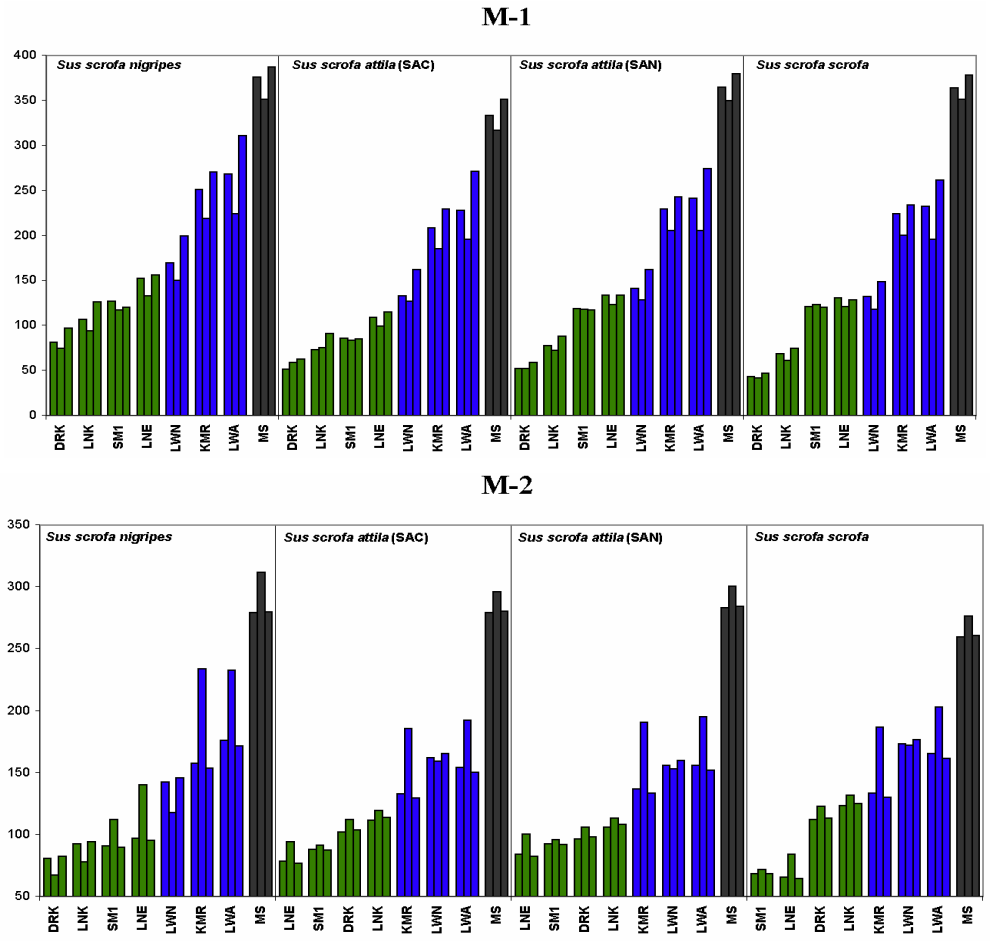



**Figure 2.** Distances between wild boar populations and domestic pig breeds according to maps constructed in two principal component coordinates. For each breed, the first bar presents Euclidean distances; the second, Harpending–Jenkins' ones; and the third, Nei's ones, respectively. Population designations follow **Table 1**.

by selection for a less intense fat deposition. The elevated fat deposition in domestic pigs in comparison with wild or early domesticated forms is an obvious abnormality, which may have been caused by the breakdown of some genes owing to PERV insertion. Therefore, the natural selection against these breaks favoured alleles characteristic of the original wild boar or similar. This may have resulted in convergent similarity between meat pig breeds and wild boars. Domestic breeds of the universal meat-and-lard morphotype should possess a certain number of loci in the genome that would determine the fat deposition degree corresponding to this morphotype. Mutations in some of these loci caused by PERV insertion give rise to the desirable trait; therefore, this morphotype diverges more from wild boars than breeds of the meat type in the frequencies of chromosomes carrying certain PERV classes and type combinations. Miniature pigs were raised by intensive selection for a smaller adult body size with minimum selection for other traits. The development of irrelevant traits should be sufficient for no more than maintenance of the population. For this reason, it is likely that the miniature pig genome was enriched in loci whose function was disrupted by PERV insertions. In some cases, this favoured the desired trait (small size), and in other cases this was of no significance, because no selection for commercial traits was conducted.

In summary, we analyzed in this study patterns of differentiation of domestic and wild pigs in the frequencies of chromosomes carrying certain PERV classes and type combinations. With regard to this differentiation, we demonstrated that the convergence processes were at least no less significant than the divergence ones. It appears that PERVs were not neutral elements in the evolution of the pig genome.

## 5. ACKNOWLEDGEMENTS

This study was supported by the *Gene pool dynamics* project of the Russ. Acad. Sci Presidium program *Biodiversity and gene pool dynamics*.

## REFERENCES

[1] Niebert, M. and Tonjes, R.R. (2005) Evolutionary spread and recombination of porcine endogenous retroviruses in suiformes. *Journal of Virology*, **79(1)**, 649-654.

[2] Akiyoshi, D.E., Denaro, M., Zhu, H., *et al*. (1998) Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. *Journal of Virology*, **72(5)**, 4503-4507.

[3] Le Tissier, P., Stoye, J. P., Takeuchi, Y., *et al*. (1997) Two sets of human-tropic pig retrovirus. *Nature*, **389(6652)**, 681-682.

[4] Takeuchi, Y., Patience, C., Magre, S., *et al*. (1998) Host range and interference studies of three classes of pig endogenous retrovirus. *Journal of Virology*, **72(12)**, 9986-9991.

[5] Bosch, S., Arnauld, C. and Jestin, A. (2000) Study of full-length porcine endogenous retrovirus genomes with envelope gene polymorphism in a specific-pathogen-free large white swine herd. *Journal of Virology*, **74(18)**, 8575-8581.

[6] Herring, C., Quinn, G., Bower, R., Parsons, N., Logan, N. A., Brawley, A., Elsome, K., Whittam, A., Fernandez-Suarez, X.M., Cunningham, D., Onions, D., Langford, G. and Scobie, L. (2001) Mapping full-length porcine endogenous retroviruses in a large white pig. *Journal of Virology*, **75(24)**, 12252-12265.

[7] Rogel-Gaillard, C., Bourgeaux, N., Billault, A., Vaiman, M. and Chardon, P. (1999) Construction of a swine BAC library: Application to the characterization and mapping of porcine type C endoviral elements. *Cytogenetics and Cell Genetics*, **85(3-4)**, 205-211.

[8] Lee, J.H., Webb, G.C., Allen, R.D. and Moran, C. (2002) Characterizing and mapping porcine endogenous retroviruses in Westran pigs. *Journal of Virology*, **76(11)**, 5548-5556.

[9] Niebert, M. and Tönjes, R.R. (2003) Analyses of prevalence and polymorphisms of six replication-competent and chromosomally assigned porcine endogenous retroviruses in individual pigs and pig subspecies. *Virology*, **313(2)**, 427-434.

[10] Edamura, K., Nasu, K., Iwami, Y., Nishimura, R., Ogawa, H., Sasaki, N. and Ohgawara, H. (2004) Prevalence of porcine endogenous retrovirus in domestic pigs in Japan and its potential infection in dogs xenotransplanted with porcine pancreatic islet cells. *Journal of Veterinary Medical Science*, **66(2)**, 129-135.

[11] Jin, H., Inoshima, Y., Wu, D., *et al*. (2000) Expression of porcine endogenous retrovirus in peripheral blood leucocytes ten different breeds. *Transplant Infectious Disease*, **2(1)**, 11-14.

[12] Zhang, L., Yu, P., Li, S.F., Bu, H., Li, Y.P., Zeng, Y.Z. and Cheng, J.Q. (2004) Phylogenetic relationship of porcine endogenous retrovirus (PERV) in Chinese pigs with some type C retrovirus. *Virus Research*, **105(2)**, 167-173.

[13] Aitnazarov, R.B., Ermolaev, V.I., Nikitin, S.V., *et al*. (2006) Associations between various endogenous virus types and genetic markers in domestic and wild pig populations. *Russian Agricultural Sciences* (*Doklady Rossiiskoi Akademii Sel'skokhozyaistvennykh Nauk*), **4**, 39-43.

[14] Nikitin, S.V., Yudin, N.S., Knyazev, S.P., *et al*. (2008) Frequency of chromosomes carrying endogenous retroviruses in the populations of domestic pig and wild boar. *Russian Journal of Genetics*, **44(6)**, 686-693.

[15] Zhivotovskii, L.A. (1991) Populational biometry. in Russian, Nauka, Moscow.

[16] Weir, B.S. (1995) Genetic data analysis. in Russian, MIR, Moscow.

[17] Harpending, H.C. and Jenkins, T. (1973) Genetic distances among Southern African Populations. In: *Methods and Theories of Anthropological Genetics*. University of New Mexico Press, Albuquerque, 177-199.

[18] Porter, V. (1993) Pigs: A handbook to the breeds of the

world. Comstoc Publishing Associates, Ithaca–New York.
[19] Gorelov, I.G. (1999) Siberian minipig, a new biomodel. *Science in Siberia*, **35(2221)**, 4-10.
[20] Knyazev, S.P., Tikhonov, V.N., Suzuki, S., *et al*. (1985) Genetic peculiarities of domestic and wild pigs of Eurasia by serum polymorhic systems. *Zoologichesky Zhournal*, **61(10)**, 1563-1568.
[21] Ovsyannikov, A.I. (1951) Raise of the Kemerovo pig breed. in Russian, Novosibirskoe Oblastnoe Gosudarstvennoe Izdatel'stvo, Novosibirsk.
[22] Gudilin, I.I., Dement'ev, V.N., *et al*. (2003) The Kemerovo pig breed. Siberian Branch of the Russian Academy of Agricultural Sciences, Novosibirsk.
[23] Knyazev, S.P., Nikitin, S.V., Kirichenko, A.V., *et al*. (2005) Differentiation of wild and domestic pig populations according to serum allotypes. *Sel'skokhozyaistvennaya Biologiya. Biologiya Zhivotnykh*, **6**, 100-105.
[24] Nikitin, S.V., Knyazev, S.P., Nikolaev, A.G., *et al*. (2006) Diversity of wild and domestic pig populations estimated by a set of serum allotypes. *Russian Journal of Genetics*, **42(3)**, 317-326.
[25] Vila, C., Seddon, J. and Ellegren, H. (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trend in Genetics*, **21(4)**, 214-218.
[26] Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., *et al*. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Science*, USA, **104(39)**, 15276-15281.

**Vol.2, No.6, 535-543 (2010)**
doi:10.4236/ns.2010.26067

**Natural Science**

# Phylogeny of γ-proteobacteria inferred from comparisons of 3' end 16S rRNA gene and 5' end 16S-23S ITS nucleotide sequences

**Sabarimatou Yakoubou[1,2], Jean-Charles Côté[1*]**

[1]Agriculture and Agri-Food Canada, Research Centre, Gouin Blvd, St-Jean-sur-Richelieu, Québec, Canada; *Corresponding Author: Jean-Charles.Cote@agr.gc.ca
[2]Département des Sciences Biologiques, Université du Québec à Montréal, CP 8888, Succ. "Centre-Ville" Montréal, Québec, Canada

Received 27 January 2010; revised 25 March 2010; accepted 12 April 2010.

## ABSTRACT

**The phylogeny of γ-proteobacteria was inferred from nucleotide sequence comparisons of a short 232 nucleotide sequence marker. A total of 64 γ-proteobacterial strains from 13 Orders, 22 families, 40 genera and 59 species were analyzed. The short 232 nucleotide sequence marker used here was a combination of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene and a 75 nucleotide sequence at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) sequence. Comparative analyses of the 3' end of the 16S rRNA gene nucleotide sequence showed that the last 157 bp were conserved among strains from same species and less conserved in more distantly related species. This 157 bp sequence was selected as the first part in the construction of our nucleotide sequence marker. A bootstrapped neighbor-joining tree based on the alignment of this 157 bp was constructed. This 157 bp could distinguish γ-proteobacterial species from different genera from same family. Closely related species could not be distinguished. Next, an alignment of the 16S-23S ITS nucleotide sequences of alleles from same bacterial strain was performed. The first 75 bp at the 5' end of the 16S-23S ITS was highly conserved at the intra-strain level. It was selected as the second part in the construction of our nucleotide sequence marker. Finally, a bootstrapped neighbor-joining tree based on the alignment of this 232 bp sequence was constructed. Based on the topology of the neighbour-joining tree, four major Groups, Group I to IV, were revealed with several sub-groups and clusters. Our results, based on the 232 bp sequence were, in general, in agreement with the phylogeny of γ-proteobacteria based on the 16S rRNA gene. The use of this 232 bp sequence as a phylogenetic marker presents several advantages over the use of the entire 16S rRNA gene or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. First, this marker is not allele-dependant. Second, this 232 bp marker contains 157 bp from the 3' end of the 16S rRNA gene and 75 bp from the 5' end of the 16S-23S ITS. The 157 bp allows discrimination among distantly related species. Owing to its higher rate of nucleotide substitutions, the 75 bp adds discriminating power among closely related species from same genus and closely related genera from same family. Because of its higher percentage of nucleotide sequence divergence than the 16S rRNA gene, the 232 bp marker can better discriminate among closely related γ-proteobacterial species. Third, the method is simple, rapid, suited to large screening programs and easily accessible to most laboratories. Fourth, this marker can also reveal γ-proteobacterial species which may appear misassigned and for which additional characterization appear warranted.**

**Keywords:** γ-Proteobacteria; 16S rRNA; 16S-23S ITS; Phylogeny

## 1. INTRODUCTION

The phylum proteobacteria or ''purple bacteria and their relatives'' encompasses bacteria with a wide variety of phenotype and physiological attributes and habitats [1,2]. Proteobacteria have been classified based on the homology of 16S ribosomal RNA or by hybridization of ribosomal DNA with 16S and 23S ribosomal RNA [3-6]. They have been subdivided in five major classes: α-, β-, γ-, δ- and ε- [7-9].

Most γ-proteobacteria are Gram-negative. This class

comprises 14 Orders and more than 200 genera. The γ-proteobacteria exhibit a wide range of metabolic diversity. Most are chemo-organotrophs, some are phototrophs or chemolitotrophs [1,2,10,11]. This class includes several medically and scientifically important bacteria. Some genera are human (*Klebsiella*, *Shigella*, *Salmonella*, *Yersina*, *Vibrio*), animal (*Pasteurella*) or plant pathogens (*Pseudomonas*, *Xanthomonas*, *Xylella*). Others are obligate endosymbionts (*Buchnera*, *Sodalis*, *Wigglesworthia* and *Coxiella*) [10-12]. Because of their biological importance, γ-proteobacteria are extensively studied.

The 16S ribosomal RNA (rRNA) gene has been established as the macromolecule of choice for phylogenetic analyses [5,13]. The current phylogeny of γ-proteobacteria is based on the homology of 16S rDNA nucleotide sequences [3-6,11,14].

The 16S-23S internal transcribed spacer (ITS) region is more variable than the 16S rRNA gene. It has been used, among others, in the study of specific γ-proteobacterial diversity at the species level, including *Escherichia*, *Haemophilus*, *Xanthomonas*, *Klebsiella* and *Pseudomonas* [15-19].

Additional approaches, based on different genes, have been used for the study of γ-proteobacterial phylogeny [12,20-25]. Very recently, Gao *et al*., [26] have used a combination of phylogenomic and comparative genomic approaches to reconstruct the phylogeny of γ-proteobacteria.

In an earlier work on the bacterial Gram-positive *Bacillus* genus and related genera [27], a short 220 bp nucleotide sequence "marker" was used to reconstruct their phylogeny. This 220 bp marker was a combination of a 150 bp sequence at the 3' end of the 16S rRNA gene and a 70 bp sequence at the 5' end of the 16S-23S ITS sequence. Owing to its higher rate of nucleotide substitution, the 70 bp sequence at the 5' end of the 16S-23S ITS sequence added a greater discriminatory power among closely related species than 16S rRNA gene nucleotide sequences alone. They showed that the phylogeny inferred from the 220 bp marker was in agreement with then current classifications based on phenetic and molecular data. The marker also indentified species which appeared misassigned. It also created new clusters suggesting the creation of new taxa levels. In a very recent study, we [28] have tested whether or not this marker could reconstruct the phylogeny of the bacterial Gram-positive Order of the *Bacillales*.

In the current study, we further assess the usefulness of a similar marker among 13 of the 14 γ-proteobacterial Orders. The last 157 bp at the 3' end of the 16S rRNA gene was combined with the first 75 bp at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) to yield a single 232 bp DNA marker. This marker was used to reconstruct the phylogeny of γ-proteobacteria. A total of 64 γ-proteobacteria from 13 Orders, 22 families, 40 genera and 59 species was analyzed.

## 2. MATERIALS AND METHODS

### 2.1. Bacterial Species and Strains

A total of 64 γ-proteobacterial species and strains were analyzed. They were selected on the basis that their complete genome sequences were freely available in GenBank, at the National Center for Biotechnology Information (NCBI) completed microbial genomes database (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/Complete.html). They encompassed 13 Orders, 22 families, 40 genera and 59 species. These 13 γ-proteobacterial Orders included six *Aeromonadales* families, one *Cardiobacteriales* family, two *Chromatiales* families, one *Enterobacteriales* family, two *Legionellales* families, one *Methylococcales* family, two *Oceanospirillales* families, one *Pasteurellales* family, two *Pseudomonadales* families, two *Thiotrichales* families, one *Vibrionales* family and one *Xanthomonadales* family. All bacterial species and strains and the GenBank accession number for their fully sequenced genome are listed in **Table 1**.

### 2.2. Sequences Analysis

The 16S rRNA gene nucleotide sequences were retrieved from GenBank (**Table 1**) for the 64 γ-proteobacteria species and strains under study. First, all 64 sequences were aligned using ClustalW [29] (data not shown). Next, the 3' end of the 16S rRNA gene nucleotide sequences of alleles from same bacterial strain, of alleles from different strains from same species, and of alleles from different species from same genus (data not shown) were aligned using ClustalW [29]. The length of the nucleotide sequence most conserved was determined at 157 bp. Likewise, the 16S-23S Internal Transcribed Spacer (ITS) nucleotide sequences of alleles from same bacterial strain were also aligned using ClustalW. The length of the nucleotide sequence most conserved was determined at 75 bp.

These two most conserved nucleotide sequences, the 157 bp at the 3' end of 16S, and the 75 bp at the 5' end of 16S-23S ITS were combined into a single 232 bp sequence for each bacterial species and strain under study. This 232 bp sequence will be used here as a phylogenetic marker for the γ-proteobacteria under study.

### 2.3. Phylogenetic trees

Two neighbor-joining trees were constructed [30], a first one based on the alignment of the last 157 bp at the 3' end of the 16S rRNA gene described above, a second

one based on the alignment of the 232 bp sequence also described above. Both trees were bootstrapped using 1,000 random samples of sites from the alignment, all using CLUSTAL W [29] at the DNA Data Bank of Japan (DDBJ) (http://clustalw.ddbj.nig.ac.jp/top-e.html), with the Kimura's parameter method [31]. The neighbor-joining tree was drawn using TreeView (version 1.6.6) [32,33].

**Table 1.** γ-proteobacteria strains used in this study.

| Orders, Families, Genera, Species | Strain/Source | GenBank Accession no. |
|---|---|---|
| *Aeromonadales* | | |
| *Aeromonodaceae* | | |
| *Aeromonas hydrophyla* | ATCC 7966 | NC_008570.1 |
| *Aeromonas salmonida* | A449 | NC_009348.1 |
| *Alteromonadales* | | |
| *Alteromonadaceae* | | |
| *Marinobacter aquaeolei* | VT8 | NC_008740.1 |
| *Saccharophagus degradans* | 2-40 | NC_007912.1 |
| *Colwelliaceae* | | |
| *Colwellia psychrerythrea* | 34H | NC_003910.7 |
| *Idiomarinaceae* | | |
| *Idiomarina ihoihiensis* | L2TR | NC_006512.1 |
| *Pseudoalteromonadaceae* | | |
| *Pseudoalteromonas atlantica* | T6c | NC_008228.1 |
| *Pseudoalteromonas haloplanktis* | TAC125 | NC_007481.1 |
| *Shewanellaceae* | | |
| *Shewanella amazonensis* | SB2B | NC_008700.1 |
| *Shewanella denitrificans* | OS217 | NC_007954.1 |
| *Shewanella frigidimarina* | NCIMB 400 | NC_008345.1 |
| *Shewanella oneidensis* | MR-1 | NC_004347.1 |
| *Cardiobacteriales* | | |
| *Cardiobacteriaceae* | | |
| *Dichelobacter nodosus* | VCS1703A | NC_009446.1 |
| *Chromatiales* | | |
| *Chromatiaceae* | | |
| *Nitrosococcus oceani* | ATCC 19707 | NC_007484.1 |
| *Ectothiorhodospiraceae* | | |
| *Alkalilimnicola ehrlichei* | MLHE-1 | NC_008453.1 |
| *Halorhodospira halophila* | SL1 | NC_008789.1 |
| *Enterobacteriales* | | |
| *Enterobacteriaceae* | | |
| *Buchnera aphidicola* | APS | NC_002528.1 |
| *Citrobacter koseri* | ATCC BAA-895 | NC_009792.1 |
| *Enterobacter sakazakii* | ATCC BAA-894 | NC_009778.1 |
| *Enterobacter sp* | 638 | NC_009436.1 |
| *Escherichia coli* | CFT073 | NC_004431.1 |
| *Escherichia coli* | O157:H7 Sakai | NC_002695.1 |
| *Klebsiella pneumoniae* | 342 | NC_011283.1 |
| *Photorhabdus luminescens* | NTUH-K2044 | NC_012731.1 |
| *Salmonella enterica* | TT01 | NC_005126.1 |
| *Salmonella enterica* | Ty2 | NC_004631.1 |
| *Salmonella enterica* | arizonae | NC_010067.1 |
| *Shigella boydii* | Sb227 | NC_007613.1 |
| *Shigella dysenteriae* | Sd197 | NC_007606.1 |
| *Shigella flexineri* | 2457T | NC_004741.1 |
| *Sodalis glossinidus* | morsitans | NC_007712.1 |
| *Wigglesworthia glossinidia* | | NC_004344.2 |
| *Yersinia enterocolitica* | 8081 | NC_008800.1 |
| *Yersinia pestis* | CO92 | NC_003143.1 |
| *Yersinia pseudotuberculosis* | IP31758 | NC_009708.1 |
| *Legionellales* | | |
| *Coxiellaceae* | | |
| *Coxiella burneti* | Dugway 5j108-111 | NC_009727.1 |
| *Coxiella burnetii* | RSA 493 | NC_002971.3 |
| *Legionellaceae* | | |
| *Legionella pneumophila* | Lens | NC_006369.1 |
| *Legionella pneumophila* | Corby | NC_009494.1 |
| *Methylococcales* | | |
| *Methylococcaceae* | | |
| *Methylococcus capsulatus* | Bath | NC_002977.6 |
| *Pseudomonadales* | | |
| *Moraxellaceae* | | |
| *Acinetobacter baumannii* | ATCC 17978 | NC_009085.1 |
| *Acinetobacter sp.* | ADP1 | NC_005966.1 |
| *Psychrobacter arcticus* | 273-4 | NC_007204.1 |
| *Psychrobacter cryohalolentis* | K5 | NC_007969.1 |
| *Pseudomonadaceae* | | |
| *Pseudomonas fluorescens* | Pf5 | NC_004129.6 |
| *Pseudomonas syringea* | DC3000 | NC_004578.1 |
| *Thiotrichales* | | |
| *Francisellaceae* | | |
| *Francisella philomiragia* | *philomiragia* | NC_010336.1 |
| *Francisella tularensis* | *horlatica* | NC_007880.1 |
| *Piscirickettsiaceae* | | |
| *Thiomicrospira crunogena* | XCL-2 | NC_007520.2 |
| *Vibrionales* | | |
| *Vibrionaceae* | | |
| *Photobacterium profundum* | SS9 | NC_006370.1 |
| *Vibrio cholerae* | N16961 | NC_002505.1 |
| *Vibrio parahaemolyticus* | RIMD 2210633 | NC_004603.1 |
| *Vibrio vulnificus* | CMCP6 | NC_004459.2 |
| *Xanthomonadales* | | |
| *Xanthomonadaceae* | | |
| *Xanthomonas axonopodis* | 306 | NC_003919.1 |
| *Xanthomonas campestris* | 8004 | NC_007086.1 |
| *Xylella fastidiosa* | 9a5c | NC_002488.3 |

## 3. RESULTS AND DISCUSSION

In a previous study [27], on the bacterial genus *Bacillus* and closely-related genera, we reported the development of a short DNA marker that could be used to reconstruct their phylogeny. This marker was a combination of the last 150 bp at the 3' end of the 16S rRNA gene and the first 70 bp at the 5' end of the 16S-23S rRNA internal transcribed spacer (ITS) into a single 220 bp "marker". It could cluster *Bacillus* species and species from closely related genera into taxa akin to genera and could also distinguish closely related species. The 3' end of the 16S rRNA gene contained three regions that were known to be highly conserved among bacteria [34]. The 5' end of the 16S-23S rRNA ITS was conserved among alleles from same strains [27].

In the current study on γ-proteobacteria, we further assessed the usefulness of this marker. The sizes of the 3' end of the 16S rRNA gene and the 5' end of the 16S-23S rRNA ITS retained here for the construction of our phylogenetic marker for γ-proteobacteria are slightly different at 157 and 75 bp, respectively, for a total marker size of 232 bp. These sizes were selected as follows: first, an alignment of the 16S rRNA gene nucleotide sequence of alleles from same strain showed that these sequences were highly conserved. The intra-strain alleles shared 99% nucleotide sequence identities. Alleles from species from same genus, however, covered a wider spectrum of nucleotide sequence identities. Whereas alleles from *Pseudomonas* (*Ps*.) *fluorescens* Pf-5 and from *Ps. syringae* pv. tomato strain DC3000 share 98% nucleotide sequence identities, alleles from *Vibrio* (*V*.) *cholerae* and from *V. parahaemolyticus* RIMD 2210633 share 92% nucleotide sequence identities. Comparative analyses of the 3' end of the 16S rRNA gene nucleotide sequence showed that the last 157 bp were in many cases highly conserved among strains from same species. This is exemplified by *Salmonella* (*Sal*.) *enterica* arizonae and *Sal. enterica* Ty2 which share 99% nucleotide sequence identities over the last 157 bp at the 3' end of the 16S rRNA gene. Species from same genus share lower nucleotide sequence identities. This is exemplified by *Shewanella* (*She*) *amazonensis* and *She. denetrificans* which share 94% nucleotide sequence identities over the last 157 bp at the 3' end of the 16S rRNA gene.

A bootstrapped neighbor-joining tree based on the alignment of this 157 bp located at the 3' end of the 16S rRNA gene was constructed (**Figure 1**). Although, in most cases, this 157 bp could distinguish species from different genera from same family, in some cases, closely related species from different genera from same family appeared undistinguishable. This is the case for *Mannheimia succiniciproducens* and both *Actinobacillus* (*Act*), species, *Act. pleuropneumonia* and *Act. succinogenes*. This is also true for *Xylella* (*Xy*.) *fastidiosa* and both *Xanthomonas* (*X*.) species, *X. axonopodis pv. citri* str. 306 and *X. campestris pv. campestris* str.8004. And this is true for the human and animal pathogenic *Enterobacteriaceae*: the *Yersinia*, *Shigella*, *Klebsiella*, *Escherichia*, *Enterobacter* and *Citrobacter* species. In some cases, this 157 bp could distinguish species from same genus as exemplified by the *Psychrobacter*, *Pseudomonas*, *Shewanella* and *Vibrio* species. In other cases, closely related species from same genus could not be distinguished as exemplified by the *Francisella*, *Aeromonas*, *Actinobacillus*, *Haemophilus*, *Acinetobacter* and *Xanthomonas* species.

In all cases, this 157 bp could distinguish species from different families, with one exception: *She. amazonensis* and *Photobacterium* (*Ph*.) *profundum*, members of the *Shewanellacea* and *Vibrionacea* family, respectively. Both appear undistinguishable. Clearly, this 157 bp seq uence cannot distinguish closely-related species. An additional DNA sequence appears necessary to better distinguish closely-related species.

Next, an alignment of the 16S-23S ITS nucleotide sequences of alleles from same bacterial strain was carried with a subset of the bacteria under study: *Xanthomonas campestris* pv. campestris str. 8004, *Ps. syringae* pv. tomato str. DC3000, *Act. succinogenes* 130Z, *E. coli* K12, *V. parahaemolyticus* RIMD 2210633 and *Shigella flexneri* 2a str. 301 (**Figure 2**). The total number of alleles vary from two to ten for *X. campestris* pv. campestris str. 8004 and *V. parahaemolyticus* RIMD 2210633, respectively. The allelic sequences were highly homologous for some species and highly hererologous for others. *Xanthomonas campestris* pv. campestris str. 8004 and *Ps. syringae* pv. tomato str. DC3000 carry two and five identical alleles, respectively. *Actinobacillus succinogenes* 130Z, *E. coli* K12, *V. parahaemolyticus* RIMD 2210633 and *Shigella flexneri* 2a str. 301 carry five, eight, ten and six alleles respectively, with varying level of heterogeneity, where highly homologous alleles are grouped together and can be distinguised from different alleles in same strain (**Figure 2**). Alleles carry from zero to four tRNA genes. An alignment of the nucleotide sequences among alleles at the intra-strain level required the introduction of several gaps. The first 75 bp at the 5' end of the 16S-23S ITS, however, was highly conserved at the intra-strain level. It was retained here for the construction of our phylogenetic marker. The two conserved nucleotide sequences identified above, the 157 bp at the 3' end of 16S rRNA gene and the 75 bp at the 5' end of 16S-23S ITS, were combined into a single 232 bp sequence. This will be used here as a phylogenetic marker for the γ-proteobacteria under study.

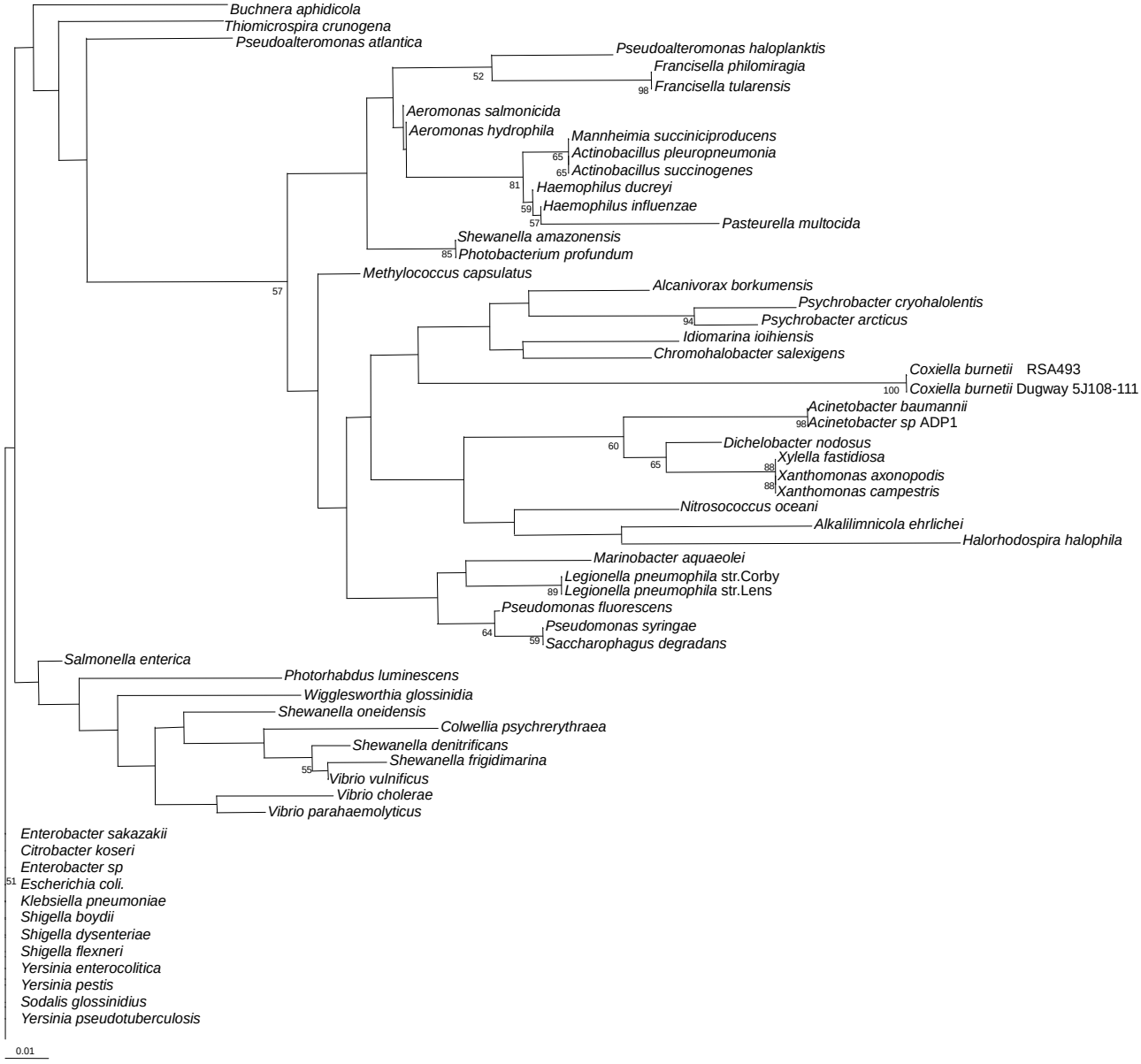



**Figure 1.** Bootstrapped neighbor-joining tree of γ –proteobacteria species inferred from the alignment of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nucleotide difference.

A bootstrapped neighbor-joining tree based on the alignment of a 232 bp sequence was constructed (**Figure 3**). Based on the topology of the neighbor-joining tree, four major Groups, Group I to IV are revealed (**Figure 3**). Based on nucleotide sequence identities, sub-groups and clusters can be formed. Group I contains seven Orders and nine families. Group II contains eight Orders and eleven families. Group III contains three Orders and five families. Group IV contains one Order and one family. Of the 13 Orders under study, four are present in more than Group one. The *Thiotrichales* are present in Groups I and III. The *Alteromonadales* are present in Groups I, II and III. The *Legionellales* are present in Groups I and II. The *Vibrionales* are present in Groups I and III. All other nine Orders are present in a single Group. Of the 22 families under study, four are present in more than one Group. The *Alteromonadaceae* are present in Groups I and II. The *Pseudoalteromonadaceae* (*Pse*) are present in Groups I and III. The *Vibrionaceae* are present in Groups I and III. The *Shewanellaceae* are present in Groups I and III. All other 18 families are present in a single Group. All species from same genus are present in

same Group with the exception of *Pse. haloplanktis* and *Pse. atlantica* present in Group I and III, respectively.

Group I can be sub-divided into six sub-groups, sub-group I-1 to I-6. Sub-group I-1 contains both species of the *Francisellaceae* family. Sub-group I-2 contains three species, members of two families: *Pseudomonadaceae* and *Alteromonadaceae*. Sub-group I-3 comprises *Pse. haloplanktis* and both *Legionella pneumophila* strains. Although in the same sub-group, *Pse. haloplanktis* shows 20% nucleotide sequence divergence with the two *Legionella* strains. Both *Legionella* strains are tightly grouped together and form cluster I-3-1. Sub-group I-4 contains *Ph. profundum* and *She. amazonensis*, two species from two different families, *Vibrionaceae* and *Shewanellaceae*, respectively. Although they appear very similar on the neighbor-joining tree, both sequences show 15% nucleotide divergence. Sub-group I-5 contains both *Aeromonas* species, tightly grouped together. Sub-group I-6 contains all six *Pasteurellaceae* species. The *Mannheimia* species is tightly grouped with the two *Actinobacillus* species and form cluster I-6-1. Both *Haemophilus* species form cluster I-6-2. In Group I, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families.

Group II can be sub-divided into four sub-groups, sub-group II-1 to II-4, and one ungrouped species, *Methylococcus capsulatus*. Sub-group II-1 comprises members of four families from four Orders. Members of sub-group II-1 share up to 30% nucleotide sequence divergences. Closely related species can be further grouped together. This is the case for the two *Acinetobacter* species and all three *Xanthomonadaceae* species which form cluster II-1-1 and II-1-2, respectively. Sub-group II-2 contains *Nitrosoccocus oceani* and *Halorhodospira halophila*, member of the *Chromatiaceae* and *Ectothiorhodospiracea* family respectively. Both species show 42% nucleotide sequence divergences. Both families belong to the *Chromatiales* Order. Sub-group II-3 contains six species, from five families and three Orders. Members of sub-group II-3 share up to 24% nucleotide sequence divergences. Two closely related species, *Psychrobacter* (*Psy*) *cryohalolentis* and *Psy. arcticus* form cluster II-3-1. Sub-group II-4 comprises both strains of *Coxiella burnetti*, members of the *Coxelliaceae* family. In Group II, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families.

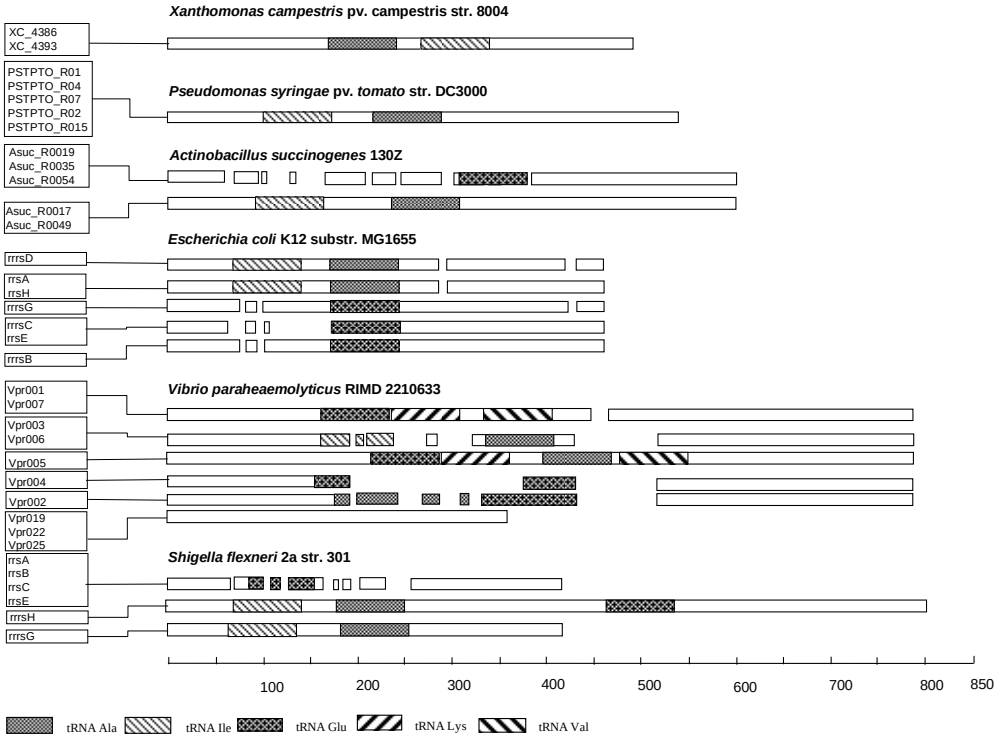



**Figure 2.** Schematic representation of allelic 16S-23S rDNA Internal Transcribed Spacer of γ –proteobacteria.The non filed boxes represent regions of homologus nucleotide sequences between allelic ITS of the same bacteria. Filed boxes represent tRNA. The blank spaces between boxes represent non conservation regions between allelic ITS of the same bacteria.

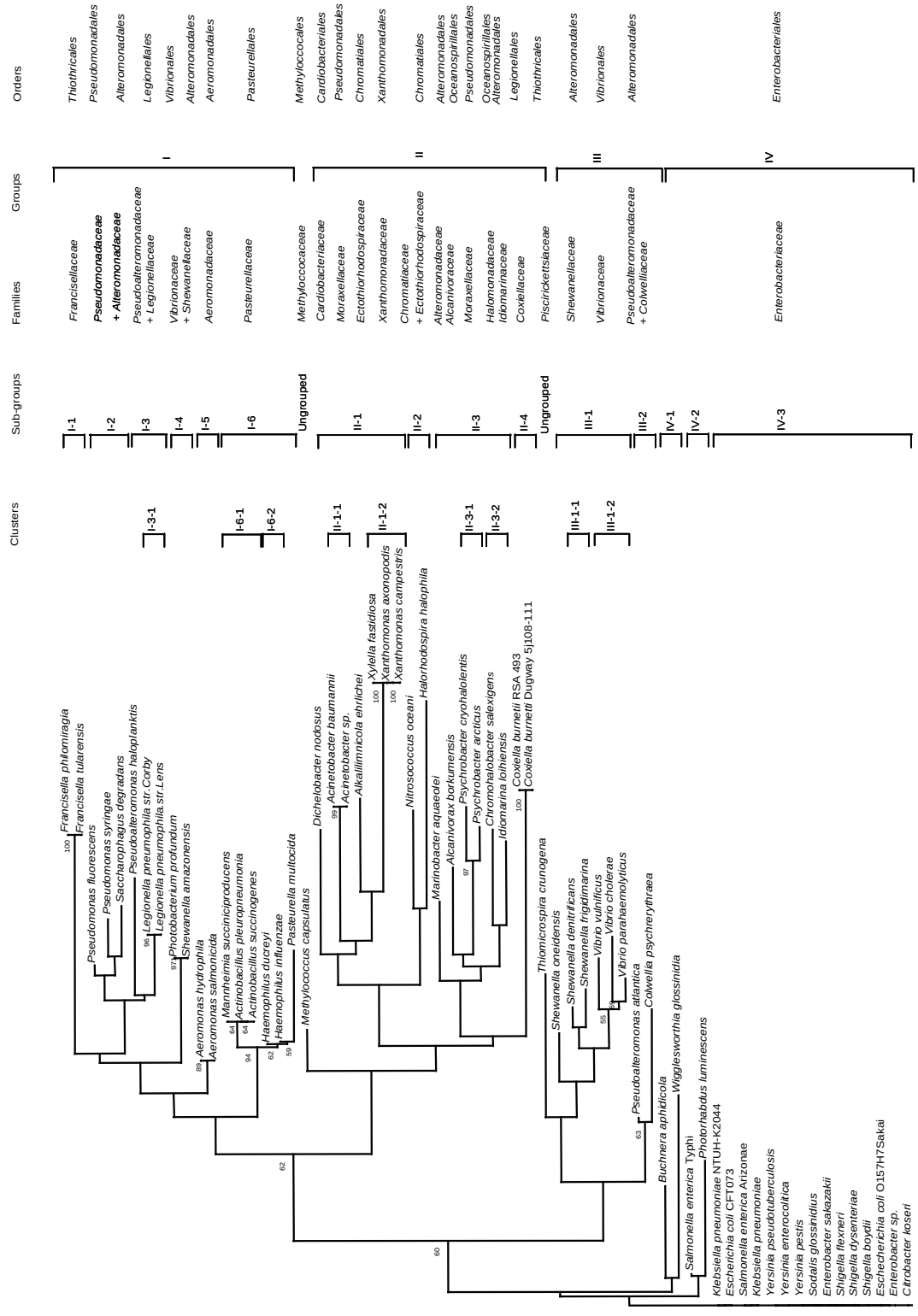



**Figure 3.** Bootstrapped neighbor-joining tree of γ-proteobacteria species inferred from the alignment of 232 nucleotide sequence marker. This 232 nucleotide sequence marker is a combination of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene and a 75 nucleotide sequence at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) sequence. Major Groups are indicated in capital roman numerals. Sub-groups and clusters are indicated in arabic numbers. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nucleotide difference.

Group III can be sub-divided into two sub-groups, sub-groups III-1 and III-2, and one ungrouped species, *Thiomicrospira crunogena*. Sub-group III-1 contains three *Shewallenaceae* and three *Vibrionaceae* species. *She. denitrificans* and *She. frigidimarina*, and the three *Vibrio* species form two clusters, III-1-1 and III-1-2, respectively. Sub-group III-2 contains two genera from two families of the same Order. Both species, on two separate branches, show 23% nucleotide sequence divergence. In Group III, both clusters comprised species from same genus. All other branches corresponded to families.

Group IV contains all the *Enterobacteriaceae* species under study. Three sub-groups can be revealed: sub-group IV-1 to IV-3. Sub-group IV-1 contains two insect obligate endosymbionts. Sub-group IV-2 contains *Salmonella enterica* Ty2 and *Photorhabtus luminescens*. All other *Enterobacteriaceae* species are in sub-group IV-3. The latter are closely related to each others.

Our results, based on the 232 bp phylogenetic marker described here are, in general, in agreement with the phylogeny of γ-proteobacteria based on the 16S rRNA gene with some exceptions. In the neighbor-joining tree, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families. As indicated above, of the 22 families under study, 18 are present in a single Group and four are present in more than one Group. These latter four families encompass the marine bacteria [35]. They are *Vibrionaceae*, in Groups I and III; and the *Alteromonas*-related protobacteria (Ivanova *et al*., 2004): *Alteromonadaceae*, in Groups I and II; *Pseudoalteromonadaceae*, in Groups I and III; and *Shewanellaceae*, in Groups I and III. Interestingly, within a Group, these marine bacteria are found in close proximity of one another. It reflects the varying level of heterogeneity among *Alteromonas*-related protobacteria. The grouping showed here is based on a 232 bp marker. The biological significance of this grouping is unknown. Clearly, however, the phylogenetic analyses of these related marine heterotrophic bacteria is a work in progress [36].

## 4. CONCLUSIONS

In conclusion, the use of this 232 bp marker presents several advantages over the use of the entire 16S rRNA gene or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. First, this marker is not allele-dependant. The 3' end of the 16S rRNA gene is highly conserved at the intra-strain level. We have shown here that although the16S-23S ITS allelic sequences can be very heterogeneous within a strain, the first 75 bp, however, are conserved among alleles from same strain in γ-proteobacteria. Clearly, any allele would generate the same results. Second, this 232 bp marker contains 157 bp from the 3' end of the 16S rRNA gene and 75 bp from the 5' end of the 16S-23S ITS. The 157 bp is highly conserved among closely related species. Owing to its higher rate of nucleotide substitutions, the 75 bp adds discriminating power among closely related species from same genus and closely related genera from same family. Because of its higher percentage of nucleotide sequence divergence than the 16S rRNA gene, the 232 bp marker can better discriminate among closely related γ-proteobacteria species. Third, the method is simple, rapid, suited to large screening programmes and easily accessible to most laboratories. More importantly, however, this 232 bp marker can group γ-proteobacteria families and genera in accordance with established phylogenies, with the exceptions indicated above. It can also reveal γ-proteobacteria species which may appear misassigned and for which additional characterization appear warranted.

## REFERENCES

[1] Stackebrandt, E., Murray, R.G.E. and Trüper, H.G. (1988) Proteobacteria classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives". *International Journal of Systematic Bacteriology*, **38(3)**, 321-325.

[2] Gupta, R.S. (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiology Reviews*, **24(4)**, 367-402.

[3] Fox, G.E., Stackebrandt, E., Hespell R.B., Gibson J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980) The phylogeny of prokaryotes. *Science*, **209(4455)**, 457-463.

[4] Woese, C.R., Stackebrandt, E., Macke, T.J. and Fox, G. E. (1985a) A phylogenetic definition of the major eubacterial taxa. *Systematic and Applied Microbiology*, **6**, 143-151.

[5] Woese, C.R. (1987) Bacterial Evolution. *FEMS Microbiology Reviews*, **51(2)**, 221-271.

[6] De Ley, J. (1992) The proteobacteria: Ribosomal RNA cistron similarities and bacterial taxonomy. In: by Balows, A., Trüper, H.G., Dworkin, M., Harder, W., and Schleifer, K.H., Eds., *The Prokaryotes*, 2nd Edition, Springer-Verlag, Berlin, 2111-2140.

[7] Woese, C.R., Stackebrandt, E., Weisburg, W.G., Paster, B. J., Madigan, M.T., Fowler, V.J., Hahn, C.M., Blanz, P., Gupta, R., Nealson, K.H. and Fox, G.E. (1984a) The phylogeny of purple bacteria: The alpha subdivision. *Systematic and Applied Microbiology*, **5**, 315-326.

[8] Woese, C.R., Weisburg, W.G., Paster, B.J., Hahn, C.M., Tanner, R.S., Krieg, N.R., Koops, H.-P., Harms, H. and Stackebrandt, E. (1984b) The phylogeny of purple bacteria: The beta subdivision. *Systematic and Applied Micro-*

*biology*, **5**, 327-336.

[9] Woese, C.R., Weisburg, W.G., Hahn, C.M., Paster, B.J., Zablen, L.B., Lewis, B.J., Macke, T.J., Ludwig, W. and Stackebrandt, E. (1985b) The phylogeny of purple bacteria: The gamma subdivision. *Systematic and Applied Microbiology*, **6**, 25-33.

[10] Brenner, D.J., Krieg, N.R., Staley, J.T. and Garrity, G.M., (Eds.) (2005) Bergey's manual of systematic bacteriology, 2nd Edition, Vol. 2 (The *Proteobacteria*), part C (The *Alpha-*, *Beta-*, *Delta-*, and *Epsilonproteobacteria*), Springer, New York.

[11] Kersters, K., Devos, P., Gillis, M., Swings, J., Vandamme, P. and Stackebrandt, E. (2006) Introduction to the *Proteobacteria*. In: Dworkin, M., Falkow, S., Rosenberg, E. Schleifer, K.H. and Stackebrandt, E., Eds., *The Prokaryotes*: *A Handbook on the Biology of Bacteria*. Springer, New York, 3-37.

[12] Belda, E., Moya, A. and Silva, F.J. (2005) Genome rearrangement distances and gene order phylogeny in γ-proteobacteria. *Molecular Biology and Evolution*, **22(6)**, 1456-1467.

[13] Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences*, USA, **87(12)**, 4576-4579.

[14] Ludwig, W. and Klenk, H.P. (2005) Overview: A phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Brenner, D.J. Krieg, N.R. Staley J.T. and Garrity, G.M., Eds., *Bergey's Manual of Systematic Bacteriology*, Springer-Verlag, Berlin, 49-65.

[15] Anton, A.I., Martinez-Murcia, A.J. and Rodriguez-Valera, F. (1998) Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *Journal of Molecular Evolution*, **47(1)**, 62-72.

[16] Giannino, V., Rappazzo, G., Scuto, A., Di Marco, O., Privitera, A., Santagati, M. and Stefani, S. (2001) rrn operons in *Haemophilus parainfluenzae* and mosaicism of conserved and species-specific sequences in the 16S-23S rDNA long spacer. *Research in Microbiology*, **152(5)**, 461-468.

[17] Goncalves, E.R. and Rosato, Y.B. (2002) Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *International Journal of Systematic and Evolutionary Microbiology*, **52(Pt 2)**, 355-361.

[18] Wang, M., Cao, B., Yu, Q., Liu, L., Gao, Q., Wang, L. and Feng, L. (2008) Analysis of the 16S-23S rRNA gene internal transcribed spacer region in *Klebsiella* species. *Journal of Clinical Microbiology*, **46(11)**, 3555-3563.

[19] Tambong, J.T., Xu, R. and Bromfield, E.S.P. (2009) Intercistronic heterogeneity of the 16S-23S rRNA spacer region among *Pseudomonas* strains isolated from subterranean seeds of hog peanut (*Amphicarpa bracteata*). *Microbiology*, **155(Pt 8)**, 2630-2640.

[20] Kunisawa, T. (2001) Gene arrangements and phylogeny in the class *Proteobacteria*. *Journal of Theoretical Biology*, **213(1)**, 9-19.

[21] Lerat, E., Daubin, V. and Moran, N.A. (2003) From gene trees to organismal phylogeny in prokaryotes: The case of the γ-proteobacteria. *Public Library of Science Biology*, **1(1)**, e19.

[22] Brown, J.R. and Volker, C. (2004) Phylogeny of γ –proteobacteria: Resolution of one branch of the universal tree? *Bioessays*, **26(5)**, 463-468.

[23] Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311 (5765)**, 1283-1287.

[24] Mrazek, J., Spormann, A.M. and Karlin, S. (2006) Genomic comparisons among γ-proteobacteria. *Environmental Microbiology*, **8(2)**, 273-288.

[25] Lee, H.Y. and Côté, J.-C. (2006) Phylogenetic analysis of γ-proteobacteria inferred from nucleotide sequence comparisons of the house-keeping genes *adk*, *aroE* and *gdh*: comparisons with phylogeny inferred from 16S rRNA gene sequences. *Journal of General and Applied Microbiology*, **52(3)**, 147-158.

[26] Gao, B., Mohan, R. and Gupta, R.S. (2009) Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria*. *International Journal of Systematic and Evolutionary Microbiology*, **59(2)**, 234-247.

[27] Xu, D. and Côté, J.-C. (2003) Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. *International Journal of Systematic and Evolutionary Microbiology*, **53(Pt 3)**, 695-704.

[28] Yakoubou, S., Xu, D. and Côté, J.-C. Phylogeny of *Bacillales* inferred from 3' 16S rDNA and 5' 16S-23S ITS nucleotide sequences. *Nature Science*, in Press.

[29] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22(22)**, 4673-4680.

[30] Saitou, N. and Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4(4)**, 406-425.

[31] Kimura, M. (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

[32] Page, R.D.M. (1996) TreeView: An application to display phylogenetic trees on personal computers. *Computer Application in the Biosciences*, **12(4)**, 357-358.

[33] Page, R.D.M. (2000) TreeView—tree drawing software for apple macintosh and windows. http://taxonomy.zoology.gla.ac.uk/rod/treeview.html

[34] Gürtler, V. and Stanisich V.A. (1996) New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology*, **142(Pt 1)**, 3-16.

[35] Suen, G., Goldman, B.S. and Welch, R.D. (2007) Predicting prokaryotic ecological niches using genome sequence analysis. *Public Library of Science Biology ONE*, **2(8)**, e743.

[36] Ivanova, E.P., Flavier, S. and Christen, R. (2004) Phylogenetic relationships among marine *Alteromonas*-like proteobacteria: Emended description of the family *Alteromonadaceae* and proposal of *Pseudoalteromonadaceae* fam. nov., *Colwelliaceae* fam. nov., *Shewanellaceae* fam. nov., *Moritellaceae* fam. nov., *Ferrimonadaceae* fam. nov., *Idiomarinaceae* fam. nov. and *Psychromona-daceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, **54(Pt 5)**, 1773-1788.

**Vol.2, No.6, 544-550 (2010)**
doi:10.4236/ns.2010.26068

**Natural Science**

# The characteristics of the chosen mycotoxins and their toxic influence on the human and animal metabolism

**Katarzyna Łazicka, Sławomir Orzechowski***

Department of Biochemistry, Faculty of Agriculture and Biology Warsaw University of Life Sciences – SGGW ul. Nowoursynowska Warsaw, Poland; *Corresponding Author: slawomir_orzechowski@sggw.pl

Received 23 December 2009; revised 10 March 2010; accepted 12 April 2010.

## ABSTRACT

**Mycotoxins are toxic secondary metabolites of fungi belonging mainly to the *Aspergillus, Penicillium* and *Fusarium* genera. They can be formed in various agricultural produce in specific conditions. These natural and zootoxic chemical compounds cause an array of diseases in people and animals, *i.e.*, mycotoxicoses. Depending on the type and dose, mycotoxins may cause liver, kidney, and lung cancers, as well as damage to the immune system, pathological changes in the nervous system, and reproduction disorders. Many mycotoxins disrupt digestion process, cause vomiting, nausea, anorexia, skin irritation and dermatitis, and even haemorrhages. A significant threat to the health of animals and people can be observed in the case of major infestation of crop ears from which foodstuff or feed are then produced. The ear infestation is facilitated by a suitable humidity and temperature during the growing season or while harvesting and storing the agricultural produce, which enhances the growth of mycotoxin-producing fungi. Suitable agricultural treatment, crop rotation, proper storing of crops in the conditions of limited humidity after the harvest as well as regular chemical analyses of the content of mycotoxins minimize their consumption and, accordingly, decrease the threat caused by these substances to the human and animal health.**

**Keywords:** Toxicology; Mycotoxins; Mycotoxicoses

## 1. INTRODUCTION

Mycotoxins are low-molecular-weight secondary metabolites produced by moulds. These metabolites constitute various groups of chemical compounds which may cause diseases or even death of an organism. However, not all chemical compounds produced by fungi are mycotoxins. The term 'mycotoxin' was first coined in 1962 after an incident in which almost 100,000 turkeys died. It was stated that the cause of such a high death rate of turkeys was the process of feeding them with peanut meal contaminated with the secondary metabolites of fungi of the *Aspergillus flavus* genus [1]. Currently, approximately 400 compounds described as mycotoxins are known, dozens of which are hazardous for the health of people and animals. Mycotoxins as chemical substances are not only difficult to be identified, but also difficult to be classified. The most important groups of mycotoxins are as follows: aflatoxins, ergot alkaloids, fumonisins, ochratoxins, trichothecenes, zearalenone and derivatives, epipolythiodioxopiperazines (ETP).

Based on the intensity and action time, the symptoms of mycotoxin poisoning are classified as acute, subacute, or chronic poisoning. Acute poisonings have a violent course and a clear reaction of the organism (e.g. the death of turkey population). On the other hand, chronic and subacute poisonings are caused by low doses of mycotoxins entering an organism for a long time, and they may cause e.g. liver cancer (aflatoxins) or kidney disorders (ochratoxins). Mycotoxins are found more often in those world regions in which less effective methods of eliminating fungi in the plant production and storing are applied (Africa, Asia, South America), as well as in significantly undernourished regions and where the inspection of food as far as the content of mycotoxins is concerned is irregular.

Mycotoxins could be divided into 2 groups: the group produced as a result of plant diseases in the vegetation season and the group of metabolites formed during improper storing of grain after the harvest. Mycotoxins have adverse effects on plants, the biological productivity of agroecosystems, the quality of surface waters, as well as on grassland and woodland ecosystems. Their harmful effect on plants is based mainly on the inhibi-

tion of germination of seeds and growth of plants, particularly the growth of roots and root hair.

## 2. THE INFLUENCE OF THE CHOSEN MYCOTOXINS ON THE HUMAN AND ANIMAL HEALTH

### 2.1. Aflatoxins

Aflatoxins are highly carcinogenic compounds as far as fish, warm-blooded animals, and people are concerned. The disease caused by these mycotoxins is called aflatoxicosis and it can lead even to death. The symptoms of chronic aflatoxicosis are liver cancer, immune system damage, as well as other ailments. Liver, which is the key organ in the mycotoxin detoxication, is the organ exposed most to the influence of one of the forms of aflatoxins, *i.e.*, aflatoxin $B_1$ ($AFB_1$). Moreover, aflatoxins may cause certain negative changes in the lungs. The development of lung cancer was observed in the workers inhaling dust contaminated with $AFB_1$ during the production of peanuts [2]. The amount of aflatoxins eaten with food is lower in the developed countries than in the developing ones. This is connected mainly with the regular food inspections being performed in order to detect aflatoxins, which in poorer countries, where food deficiency is present, are very rare or are not carried out at all. The high consumption of aflatoxins with food contributes to the development of primary liver cancer 2 to 5 times more often than in the developed countries [3].

Dogs are particularly sensitive to aflatoxins, as even relatively low doses of these mycotoxins in feed (0.5-1 mg per 1 kg b.m.) may cause their death. At doses ranging from 0.05-0.3 mg/kg b.m. of aflatoxins, provided for 6 to 8 weeks, vomiting, depression, hepatitis, coma, and jaundice are observed in dogs. Dogs are exposed to the influence of aflatoxins most often due to the introduction of grain contaminated with aflatoxins into feed production, which for this reason cannot be used for the production of food for people [4].

The extent and intensity of poisoning with aflatoxins depend on age, sex, body mass, diet, resistance to infections, presence of other mycotoxins, as well as on pharmacological agents in diet. Due to the differences in sensitivity to aflatoxins in tests performed on animals, it is difficult to explain the influence of aflatoxins on the human organism. Mycotoxin $AFB_1$ becomes active under the influence of cytochrome P-450, present in liver microsomes, in the presence of molecular oxygen and NADPH. On the other hand, the active form of $AFB_1$ may bond with DNA or proteins [5]. The $AFB_1$-DNA adducts cause the transversion mutation GC: TA.

Aflatoxins are found in the marketed animal feeds all over the world, particularly in the North and South Americas. Birdseed was the most contaminated feed among all the types of products for domestic animals probably due to the basic materials of this feed, which are grains and hazelnuts. As much as 25% of birdseed samples were contaminated with over 100 μg $AFB_1$/kg, which indicates the potential threat to the health of birds. Feeds for dogs and cats contain small amounts of aflatoxins. Almost all the feed samples that these mycotoxins were found in, contained less than 20 μg $AFB_1$/kg [6]. Such a low level of aflatoxins in the feed does not cause any symptoms of acute poisoning. However, the possible future influence of these mycotoxins on the animal metabolism should not be ignored.

### 2.2. Ergot Alkaloids

The chemical compounds belonging to this group of mycotoxins are produced as a mixture of alkaloids in sclerotia of the strains of *Claviceps* – *Claviceps purpurea*, which are common pathogens of various grass species ears. The consumption of ergot or its spores was found to be connected with the occurrence of diseases even in the antiquity. In people, the disease known as ergotism occurs after the consumption of products made of grain containing significant amounts of sclerotia, usually as bakery products. Usually two types of ergotism are diagnosed, *i.e.*, gangrenous and convulsive (Saint Anthony's Fire). The gangrenous type ergotism causes significant ischaemia of nose, ears, fingers and toes, which leads to dry gangrene, whereas the convulsive ergotism shows itself in the changes in behaviour, hypersensitivity, convulsions, and muscle tremor [7].

Penitrem A, which belongs to the group of ergot alkaloids, affects the functioning of the central nervous system probably by the interaction with neurotransmitters such as e.g. γ-aminobutyric acid (GABA) [8].

In animals the clinical symptoms of ergotism are as follows: dry gangrenes, ulcerations, necroses, blindness, convulsions, hypersensitivity and ataxia, and lactation inhibition [6]. A number of ergot alkaloids cause contractions of smooth muscles (e.g. uterus, blood vessels). Among farm animals, sheep, pigs, poultry, and cattle are exposed the most to the influence of these mycotoxins.

### 2.3. Fumonisins

Fumonisins occur in the natural environment in at least three forms: B1, B2, and B3. They are found mainly in feed and food produced from maize [9]. Their influence on the living organisms is diverse; they may cause the development of cancer in humans and animals together with other mycotoxins (intensifying their negative action) [10,11]. In horses and rabbits, fumonisins cause leukoencephalomalacia [12,13], and they disrupt sphingol-

ipid metabolism by the inhibition of (dihydro)ceramide synthesis, causing the accumulation of sphinganine in tissues, serum, and urine. The sphinganine accumulation is responsible for the most of adverse effects of these mycotoxins. The decrease in the sphingolipid complex has a negative effect on the functions of a part of the membrane proteins, such as e.g. folic acid transmitter [14] and contributes to the occurrence of nervous system disease symptoms [15].

## 2.4. Ochratoxin

Among the toxins produced by the *Aspergillus* and *Penicillum* species, ochratoxin A (OTA) has a significant meaning from the toxicological and economic points of view. This mycotoxin is formed mainly as a result of improper storing of grain. The most important adverse changes caused by the consumption of food contaminated with ochratoxin A are: kidney and liver damage and immunosuppressive action [16]. Ochratoxin A at the level of cellular metabolism affects the enzymes connected with the metabolism of phenylalanine and causes the inhibition of the mitochondrial synthesis of ATP [17] and acceleration of the peroxidation of lipids [18]. Due to the constitution of ochratoxin A (halogen derivative of phenol), the action of this mycotoxin may also result in the modification of bases constituting the nucleic acids, which may cause mutations and contribute to the development of carcinogenic diseases in the organisms exposed to its influence [19].

## 2.5. Trichothecenes

The name 'trichothecenes' includes almost 200 various compounds characterised as mycotoxins, synthesised mainly by fungi from the *Fusarium* family found on infested ears. The most important and the most commonly occurring representatives of this group of mycotoxins are deoxynivalenol (DON, vomitoxin), nivalenol (NIV), diacetoxyscirpenol (DAS), and T-2. The typical disease symptoms after the consumption of trichothecenes are: vomiting, lack of appetite, diarrhoea, haemorrhages, the nervous and immune systems disorders, and in case of direct contact dermatitis is observed. Trichothecenes are exceedingly strong inhibitors of the protein synthesis in eukaryotes. They inhibit all its stages: initiation, elongation, and termination. Trichodermin was identified as the first metabolite of the trichothecenes group and its influence on the metabolism consists in the inhibition of peptidyl transferase [20]. Trichodermin inhibits peptidyl transferase through competing with it for the binding site on a ribosome [21].

Among the naturally occurring trichothecenes, T-2 and diacetoxyscirpenol seem to be the most significant ones in the studies on their toxicity in animals. The cytotoxic influence of these mycotoxins on organisms is based on their negative influence on the immune system, manifesting itself in the decreased resistance to infection with other pathogenic microorganisms. As the final effect, a wide range of gastrointestinal, dermatological and neurological symptoms are observed after the consumption of these mycotoxins by animals [22]. In people, T-2 and diacetoxyscirpenol cause alimentary aleukia. The symptoms of this disease are: dermatitis, vomiting, and hematopoietic tissue disorders. The acute phase results in oral cavity necrosis, nosebleed, vaginal bleeding, and the nervous system disorders. Trichothecene mycotoxins induce also asthma and allergenicity [23].

## 2.6. Zearalenone

The biological activity of zearalenone (ZEA) metabolites is high. It can be generally stated that they act oestrogen-like, though their actual toxicity is low. The $LD_{50}$ of zearalenone in rats is 10 g/kg, whereas in guinea pigs it is 5 g/kg. For this reason, the more appropriate name than mycotoxin for this group of compounds is phytoestrogen (or mycoestrogen). Zearalenone or its derivatives have been patented as oral contraceptives for women, or a drug preventing postmenopausal osteoporosis [24]. The link between the consumption of mouldy grain and hyperestrogenism in pigs has been observed since 1920. After the administration *per os* of even minor doses of zearalenone to the sexually immature sows, oestrus-like symptoms without the tolerance reaction and the increase in the mass of certain parts of the reproductive system of the examined sows were observed [25]. In dogs, the studied concentrations of ZEA caused disruptions in reproduction, miscarriages, and testicular atrophy and pathological changes in the reproductive system in males [26,27]. Based on the results of epidemiological studies performed in Canada and Scandinavia, it was stated that the risk for human health after the consumption of zearalenone is not high. The recommended safe intake of zearalenone is estimated at the level of 0.05 μg/kg b.m. daily. The level of zearalenone in consumables is not checked. The presence of ZEA in diet may cause various interactions with metabolic processes in the organism, and its presence in diet together with other important mycotoxins, such as fumonisins or trichothecenes, is potentially a pathogenic factor. The manifestations of this action may be allergies or the influence on the immune system of animals and people [28].

## 2.7. Epipolythiodioxopiperazine ETP

Characteristic features of the chemical structure of mycotoxins of the ETP group are the presence of disulphide bridge that largely conditions their high reactivity and adverse influence, and the presence of at least 1 aromatic

acid radical [29]. The most thoroughly characterized mycotoxin from this group is gliotoxin, which is at the same time a drug used e.g. in the treatment of non-alcoholic fatty liver disease. Its action is based on the inhibition of the activity of the nuclear factor κB (NF-κB) and activation of apoptosis through the inhibition of mitochondrial ATP synthesis, which leads to the hyperpolarisation of mitochondrial membranes. Apoptosis was provoked by gliotoxin in various cells of the immune system (macrophages, lymphocytes T) [30,31]. Gliotoxin inhibits also the activity of reverse transcriptase, which may be used in the treatment of viral infections in people and animals [32].

Due to its structure, gliotoxin as well as the remaining mycotoxins of the ETP group may take part in the oxidation and reduction reactions in cells, which may cause the formation of reactive oxygen species (ROS) and is, beside the creation of disulfide bridges with proteins, the second main cause of a significant toxicity of this group of mycotoxins observed in animals and people. **Table 1** presents the groups of mycotoxins which are the most harmful to the organism, together with the chosen disease symptoms they cause.

## 3. METHODS OF LIMITING THE POISONING WITH MYCOTOXINS

The production of mycotoxins by various species of fungi occurring on agricultural produce happens frequently. These fungi may grow on various plants and agricultural materials at an appropriate humidity and temperature. In the case of agricultural produce, the level of contamination changes by the year, depending on the weather and other environmental factors. For example, aflatoxin is us ually hazardous during dry years, as plants are then weaker and become more susceptible to damage caused by insects and other mechanical injuries. Therefore, it is important to use a proper cultivation technology in such cases: proper fertilization, watering, gathering the crop residue from the field (mainly from crops and monocottyledon weeds), and using fungicides (usually only seed dressing before seeding is performed in the conventional and integrated agriculture). A significantly increased level of mycotoxins in agricultural produce is observed in the case of using non-plough methods of cultivation and direct seeding in comparison to plough cultivation, although in favourable weather conditions this effect is not very visible. Mycotoxins in agricultural produce may appear only after noticing the growth of mould, however not all moulds produce mycotoxins and the production does not take place in all atmospheric conditions. Therefore, an important element limiting the contamination with mycotoxins is constant monitoring of agricultural produce first and then of the agricultural materials and food in order to eliminate excessively contaminated material from subsequent stages of production of food and animal feed.

Depending on the level of contamination with mycotoxins, agricultural produce may be:

1) completely destroyed (significant contamination),

2) destined for the production of feed for animals (contamination threatens human health and life),

3) passed as fit for the production of food for people (trace amount of mycotoxins or their lack).

**Table 1.** The list of adverse effects of the chosen mycotoxins.

| Name | Toxicity class to IARC* | Symptoms/ disease entity |
|---|---|---|
| **Aflatoxins** | I | aflatoxicosis, primary liver cancer, lung neoplasm, lung cancer, failure of the immune system, vomiting, depression, hepatitis, anorexia, jaundice, vascular coagulation |
| **Ergot alkaloids** | - | ergotism, gangrene, abortion, convulsions, lactation suppression, hypersensitivity, ataxia, smooth muscle contractions, extensive saliva production, vomiting |
| **Fumonisins** | II B | diseases of the nervous system, cerebral softening, pulmonary oedema, liver cancers, kidney diseases, oesophagus cancers, anorexia, depression, ataxia, blindness, hysteria, vomiting, hypotension |
| **Ochratoxins** | II B | renal diseases, nephropathy, anorexia, vomiting, intestinal haemorrhage, tonsillitis, dehydration |
| **Trichothecenes** | - | nausea, vomiting, haemorrhages, anorexia, alimentary toxic aleukia, failure of the immune system, infants' lung bleeding, increased thirst, skin rash |
| **Zearalenone** | - | reproduction disruptions, abortions, pathological changes in the reproductive system |
| **ETP** | - | failure of the immune system, susceptibility to fungal infections, facial eczema, liver diseases, skin and nail necroses |

* IARC – International Agency for Research on Cancer

Moreover, it has to be kept in mind that giving contaminated feed to susceptible animals may lead to the decrease in the speed of their growth, development of diseases, or even to death. Foodstuffs produced from meat or milk of animals that were previously fed with feed contaminated with mycotoxins, may contain cumulated mycotoxins, e.g. ochratoxin A present in the feed for pigs is accumulated in the tissue of fattening pigs [33]. Furthermore, biotransformations of mycotoxins are also frequently observed in animals fed with the feed contaminated with mycotoxins, e.g. aflatoxins $AFB_1$ eaten by cows with feed undergo biochemical reactions to aflatoxin $M_1$ that is then found in milk.

People who eat properly, avoiding foodstuffs contaminated with moulds, are not exposed to mycotoxins. The symptoms caused by the consumption of high amount of mycotoxins are seldom found in the developed countries. Currently used methods of storing agricultural produce and their processing, as well as production of food aiming at the maximum limitation of development of moulds and inactivation or bounding of the already existing mycotoxins have contributed to this fact. Actions that contribute to the limitation of mould development at various stages of food production are based on: preservation or pasteurization of food, and roasting of grains at a high temperature, which unfortunately may have adverse nutritional effects, *i.e.*, the formation of stable free radicals. On a small scale ozonization is used, which by inhibiting the development of moulds decreases the level of e.g. aflatoxins and trichothecenes in food materials [34]. The limitation of the development of moulds is possible also by adding preservatives and bioactive substances such as e.g. benzoic, acetic, or propionic acids [35,36] to foodstuffs and feeds. On the other hand, the process of lactic fermentation is an example of the use of microorganisms to decrease the level of mycotoxins in feeds and foodstuffs. During the production of silages, the lactic fermentation bacteria contribute to the biotransformation of mycotoxins into metabolites which are not harmful to animals, and the lowered pH successively inhibits the development of mould spores. The decrease in the amount of mycotoxins and in the extent of infestation with moulds in the agricultural produce may be achieved with the use of simple methods, even by sifting grains, since screenings (damaged and poorly developed caryopses) are often more infested with moulds and hence potentially contaminated with mycotoxins to a higher degree than the fully developed grains [37]. Positive effects in the reduction of the chosen mycotoxins are produced by rinsing grains, e.g. with 1 M solution of sodium carbonate and distilled water, since often the infestation with moulds and the mycotoxins themselves are localized on the grain surface [38,39]. However, such treatment of grains requires drying at further stages of their processing, which puts the producers of feeds and foodstuffs at risk of making additional costs.

Feed and foodstuff producers often use bioactive substances such as: antioxidants (vitamin A, C, E, selenium) [40,41], chosen amino acids (leucine, isoleucine, valine, tyrosine, phenylalanine) [42], and polyunsaturated fatty acids [15,43] as diet supplements. Apart from increasing the nutritive value and safety of feeds and foodstuffs by completing the diet with vitamins, exogenous amino acids, or necessary unsaturated fatty acids, these substances minimize the adverse effects of the chosen mycotoxins (e.g. T-2) on the human and animal organism.

## 4. SUMMARY

In terms of their chemical composition, mycotoxins belong to compounds with a diversified constitution and properties. They are produced by numerous species of moulds, which makes it difficult to detect them in food. The most important groups of mycotoxins are as follows: aflatoxins, ergot alkaloids, fumonisins, ochratoxins, trichothecenes, zearalenones, and epipolythiodioxopiperazines. They have diverse toxic effects on animals and cause among others cancers, protein synthesis inhibition, failure of the immune system, skin and mucosa irritation, as well as other disorders. Usually mycotoxins enter an organism *per os* due to the consumption of food contaminated with them, through the respiratory system, or by direct contact with the skin. In the developing countries, the contamination of food and animal feed with mycotoxins is a significant problem. In the industrialized countries, appropriate techniques meant for harvesting and storing of agricultural produce and for the production of food and animal feed have been developed, which has significantly limited the occurrence of mycotoxicoses in people and animals A considerable problem to be solved is the unification of regulations and standards concerning the contents of certain mycotoxins all over the world, as well as the elaboration of cheap and reliable diagnostic tests for monitoring the contamination of agricultural produce with mycotoxins at all stages of the food production chain.

## REFERENCES

[1] Forgacs, J. (1962) Mycotoxicoses—the neglected diseases. *Feedstuffs*, **34(18)**, 124-134.

[2] Hayes, R.B., van Nienwenhuise, J.P., Raatgever, J.W. and Ten Kate, F.J.W. (1984) Aflatoxin exposure in the industrial setting: An epidemiological study of mortality. *Food and Chemical Toxicologyi*, **22(1)**, 39-43.

[3] Henry, S.H., Bosch, F.X. and Bowers, J.C. (2002) Aflatoxin, hepatitis and worldwide liver cancer risks. In: DeVries, J.W., Trucksess, M.W. and Jackson, L.S., Eds., *Mycotoxins and Food Safety*. Kluwer Academic/Plenum Publications, New York, 229-320.

[4] Stenske, K.A., Smith, J.R., Newman, S.J., Newman, L.B. and Kirk, C.A. (2006) Aflatoxicosis in dogs and dealing with suspected contaminated commercial foods. *Journal of the American Veterinary Medical Association*, **228(11)**, 1686-1691.

[5] Eaton, D.L. and Gallagher, E.P. (1994) Mechanisms of aflatoxin carcinogenesis. *Annual Review of Pharmacology and Toxicology*, **34**, 135-172.

[6] Maia, P.P. and Pereira Bastos de Siqueira, M.E. (2002) Occurrence of aflatoxins B1, B2, G1 and G2 in some Brazilian pet foods. *Food Additives and Contaminants*, **19(12)**, 1180-1183.

[7] Bennett, J.W. and Bentley, R. (1999) Pride and prejudice: The story of ergot. *Perspectives in Biology and Medicine*, **42(3)**, 333-355.

[8] Puschner, B. (2002) Mycotoxins. *Veterinary Clinics*: *Small Animal Practice*, **32(2)**, 409-419.

[9] Tekiela, A. (2007) Różnice w stopniu infekcji ziarna kukurydzy przez grzyby *Fusarium* spp. a skażenia przez mikotoksyny w uprawie mieszańców kukurydzy modyfikowanych genetycznie i ich form wyjściowych w warunkach polskich. *Kosmos*, **56(3-4)**, 301-305.

[10] Peraica, M., Radic, B., Lucic, A. and Pavlovic, M. (1990) Toxic effects of mycotoxins in humans. *Bulletin of the World Health Organization*, **77(9)**, 754-766.

[11] Pozzi, C.R., Correa, B., Xavier, J.G., Direito, G.M., Orsi, R.B. and Matarazzo, S.V. (2001) Effects of prolonged oral administration of fumonisin B1 and aflatoxin B1 in rats. *Mycopathologia*, **151(1)**, 21-27.

[12] Bucci, T., Hansen, D.K. and LaBorde, J.B. (1996) Leucoencephalomalacia and hemorrhage in the brain of rabbits gavaged with mycotoxin fumonisin B1. *Journal of Natural Toxins*, **4(1)**, 51-52.

[13] Marasas, W.F.O., Kellerman, T.S., Gelderblom, W.C.A., Coetzer, J.A.W., Thiel, P.G. and van Der Lugt, J.J. (1988) Leukoencephalomalacia in horse induced by fumonisin B1 isolated from *Fusarium moniliforme*. *The Onderstepoort Journal of Veterinary Research*, **55(4)**, 197-203.

[14] Stevens, V.L. and Tang, J. (1997) Fumonisin $B_1$-induced sphingolipid depletion inhibits vitamin uptake via the glycosylphosphatidylinositol-anchored folate receptor. *Journal of Biological Chemistry*, **272(29)**, 18020-18025.

[15] Sadler, T.W., Stevens, V.L., Merrill, A.H., Sullards, M.C., Wang, E. and Wang, P. (2002) Prevention of fumonisin $B_1$–incuced neural tube defects by folic acid. *Teratology* **66(4)**, 169-176.

[16] Kuiper-Goodman, T. and Scott, P.M. (1989) Risk assessment of the mycotoxin ochratoxin A. *Biomedical and Environmental Sciences*, **2(3)**, 179-248.

[17] Meisner, H. and Meisner, P. (1981) Ochratoxin A, an *in vivo* inhibitor of renal phosphoenolpyruvate carboxylase. *Archives of Biochemistry and Biophysics*, **208(1)**, 146-151.

[18] Rahimtula, A.D., Bereziat, J.C., Bussacchini-Griot, V. and Bartsch, H. (1988) Lipid peroxidation as a possible cause of ochratoxin A toxicity. *Biochemical Pharmacology*, **37(23)**, 4469-4475.

[19] Dai, J., Wright, M.W. and Manderville, R.A. (2003) Ochratoxin a forms a carbon-bonded c8-deoxyguanosine nucleoside adduct: Implications for c8 reactivity by a phenolic radical. *Journal of the American Chemical Society*, **125(3)**, 3716-3717.

[20] Wei, C.M., Campbell, I.M., McLaughlin, C.S. and Vaughn, M.H. (1974) Binding of trichodermin to mammalian ribosomes and its inhibition by other 12,13-epoxytrichothecenes. *Molecular and Cellular Biochemistry*, **3(3)**, 215-219.

[21] Feinberg, B. and McLaughlin, C.S. (1989) Biochemical mechanism of action of trichothecene mycotoxins. In: Beasley, V.R., Ed., *Trichothecene mycotoxicoses*: *Pathoph-Ysiologic Effects*, CRC Press, Boca Raton, **1**, 27-35.

[22] Rotter, B.A., Prelusky, D.B. and Pestka, J.J. (1996) Toxicology of deoxynivalenol (vomitoxin). *Journal of Toxicology and Environmental Health*, **48(1)**, 1-34.

[23] Pestka, J.J., Yike, I., Dearborn, D.G., Ward, M.D.W. and Harkema, J.R. (2008) Stachybotrys chartarum, trichothecene mycotoxins, and damp building—related illness: New insights into a public health enigma. *Toxicological Sciences*, **104(1)**, 4-26.

[24] Hidy, P.H., Baldwin, R.S., Greasham, R.L., Keith, C.L. and Mc-Mullen, J.R. (1977) Zearalenone and some derivatives: production and biological activities. *Advances in Applied Microbiology*, **22**, 59-82.

[25] Gajęcki, M. (2002) Zearalenone—undesirable substancje in feed. *Polish Journal of Veterinary Sciences*, **5(2)**, 117-122.

[26] Gajęcka, M., Jakimiuk, E., Polak, M., Otrocka-Domagala, I., Janowski, T., Zwierzchowski, W., Obremski, K., Zielonka, L., Apoznanski, J. and Gajęcki, M. (2004a) Zearalenone applied per os provides adverse effects in structure of chosen parts of bitch reproductive system. *Polish Journal of Veterinary Sciences*, **7(1)**, 59-66.

[27] Gajęcka, M., Jakimiuk, E., Skorska-Wyszynska, E., Zielonka, L., Polak, M., Paluszewski, A., Rybarczyk, L. and Gajęcki, M. (2004b) Influence of zearalenone mycotoxicosis on selected immunological, haematological and biochemical indexes of blood plasma in bitches. *Polish Journal of Veterinary Sciences*, **7(3)**, 175-180.

[28] Gajęcki, M., Gajęcka, M., Zielonka, Ł., Jakimiuk, E. and Obremski, K. (2006) Zearalenone as a potential allergen in the alimentary tract—a review. *Polish Journal of Food and Nutrition Sciences*, **15(3)**, 263-268.

[29] Gardiner, D. Waring, P. & Howlett, B. (2005) The epipolythiodioxopiperazine (ETP) class of fungal toxins: Distribution, mode of action, functions and biosynthesis. *Microbiology*, **151(Pt 4)**, 1021-1032.

[30] Pahl, H.L., Krauss, B., Schulze-Osthoff, K., Decker, T.E., Traenckner, B.-M., Vogt, M., Myersfl, Ch., Parksfl, T., Warring, P., Miihlbacher, A., Czernilofiky, A.P. and Baeuerle, P.A. (1996) The immunosuppressive fungal metabolite gliotoxin specifically inhibits transcription factor NF-kB. *Journal of Experimental Medicine*, **183(4)**, 1829-1840.

[31] Wright, M., Issa, R., Smart, D., Trim, N., Murray, G., Primrose, J., Arthur, M., Iredale, J. and Mann, D. (2001) Gliotoxin stimulates the apoptosis of human and rat hepatic stellate cells and enhances the resolution of liver fibrosis in rats. *Gastroenterology*, **121(3)**, 685-698.

[32] Rodriguez, P.L. and Carrasco, L. (1992) Gliotoxin: Inhibitor of poliovirus RNA synthesis that blocks the viral RNA polymerase 3Dpol. *Journal of Virology*, **66(4)**, 1971-1976.

[33] Stoev, S.D., Paskalev, M., MacDonald, S. and Mantle, P.G. (2002) Experimental one year ochratoxin A toxicosis in pigs. *Experimental and Toxicologic Pathology*, **53(6)**, 481-487.

[34] Proctor, A.D., Ahmedna, M., Kumar, J.V. and Goktepe, I. (2004) Degradation of aflatoxins in peanut kernels/flour by gaseous ozonation and mild heat treatment. *Food Additives and Contaminants*, **21(8)**, 786-793.

[35] Hazan, R., Levine, A. and Abeliovich, H. (2004) Benzoic acid, a weak organic acid food preservative, exerts specific effects on intracellular membrane trafficking pathways in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiologyi*, **70(8)**, 4449-4457.

[36] Lambert, R.J. and Stratford, M. (1999) Weak-acid preservatives: Modelling microbial inhibition and response. *Journal of Applied Microbiology*, **86(1)**, 157-164.

[37] Leung, M.C.K., Diaz-Llano, G. and Smith, T.K. (2006) Mycotoxins in pet food: A review on worldwide prevalence and preventative strategies. *Journal of Agricultural and Food Chemistry*, **54(26)**, 9623-9635.

[38] Trenholm, H.L., Charmley, L.L., Prelusky, D.B. and Warner, R.M. (1992) Washing procedures using water or sodium carbonate solutions for the decontamination of three cereals contaminated with deoxynivalenol and zearalenone. *Journal of Agricultural and Food Chemistry*, **40(11)**, 2147-2151.

[39] Trenholm, H.L., Charmley, L.L., Prelusky, D.B. and Warner, R.M. (1991) Two physical methods for the decontamination of four cereals contaminated with deoxynivalenol and zearalenone. *Journal of Agricultural and Food Chemistry*, **39(2)**, 356-360.

[40] Atroshi, F., Rizzo, A., Biese, I., Lindberg, L.A. and Saloniemi, H. (2002a) Effects of feeding T-2 toxin and deoxynivalenol on DNA and GSH contents of brain and spleen of rats supplemented with vitamin E and C and selenium combination. *Journal of Animal Physiology and Animal Nutrition*, **180(74)**, 151-167.

[41] Atroshi, F., Rizzo, A., Westermarck, T. and Ali-Vehmas, T. (2002b) Antioxidant nutrients and mycotoxins. *Toxicology*, **180(2)**, 151-167.

[42] Cavan, K.R., MacDonald, E.J. and Smith, T.K. (1988) Potential for dietary amino acid precursors of neurotransmitters to overcome neurochemical changes in acute T-2 toxicosis in rats. *Journal of Nutrition*, **118(7)**, 901-907.

[43] Pestka, J.J., Zhou, H.R., Jia, Q. and Timmer, A.M. (2002) Dietary fish oil suppresses experimental immunoglobulin a nephropathy in mice. *Journal of Nutrition*, **132(2)**, 261-269.

**Vol.2, No.6, 551-556 (2010)**
doi:10.4236/ns.2010.26069

**Natural Science**

# Effect of prolonged intake of iron enriched diet on testicular functions of experimental rats

**Mohamed M. El-Seweidy*, Mervat E. Asker, Sousou I. Ali, Hebatallah H. Atteia**

Biochemistry Department, Faculty of Pharmacy, Zagazig University; *Corresponding Author: mmelseweidy@yahoo.com

Received 28 December 2009; revised 8 January 2010; accepted 23 March 2010.

## ABSTRACT

**Iron deficiency anemia represents a common nutritional problem which affects many societies allover the world and iron fortified diet has been suggested as one of possible tools to combat and solve such problem. Present study was designed to illustrate the effect of dietary iron intake on certain biochemical markers dealing with oxidative stress, inflammatory response and cellular alterations of testicular tissues. Adult male rats which were fed on biscuits fortified with iron (0.3% ferrous sulfate) daily for 10 weeks (iron group) showed increased serum iron, ferritin, tumor necrosis factor-alpha (TNF-α), nitric oxide (NO) and decreased Testosterone level ($p < 0.05$). Testicular tissues content of Malondialdehyde (MDA), hydroxyproline (Hyp), iron showed significant increase ($p < 0.05$) and decreased glutathione (GSH) as compared to control group. Testicular tissues demonstrated massive iron distribution in sertoli interstial tissues and degeneration of germinal epithelial cells. Apparent reduction in number of sperms and spermatogenic cells were also observed. These symptoms may demonstrate that prolonged intake of Biscuit fortified with iron causes certain testicular damage through certain mechanism.**

**Keywords:** Iron Overload; Oxidative Stress; Inflammatory Response of Rat Testis

## 1. INTRODUCTION

Iron is an essential element for many cellular activities like oxygen transport, electron transfer [1] and gene regulation. However, an increase in cellular iron content may be toxic due to generation of reactive oxygen species (ROS) that damage proteins, lipids and DNA [2,3].

Iron deficiency anemia represents a common nutritional problem affecting large sector of individuals allover the world. Accordingly iron fortified diet was considered by some nutritionists as an effective tool and long term strategy to combat such kind of anemia.

This program was adopted and applied during last decades in different countries using different bakeries and breads fortified with Iron. Objectors to these idea stated that the possibility of toxicity from excess iron absorption of fortified diet was too great and thus such a measure should not be sponsored.Ten years ago Egyptian ministry of education, through its distributed institutes had adopted iron fortification program for different sectors of students especially young ages. This was achieved through daily supplementation to students of iron fortified biscuits (0.3% w/w). Till now this program was not evaluated properly, taken in consideration that additive iron here is greatly high. Healthy individuals may however face iron storage state even with moderate fortification of diet with iron, inducing inturn many varieties of chronic diseases [4]. Continuous iron intake may contribute to generation of free radicals and oxidation of cell components, an example is hepatocellular damage [5].

Testicular functions like liver cells are particularly vulnerable to such kind of injury and mostly mediated by reactive oxygen species (ROS) in consequence to iron overload. Oxidative damage here can either affect sperm cells or influence spermatogenic process which could change sperm functions. ROS can also activate transcription factors as nuclear factor-kappa B (NF-kB) which upregulates the transcription of adhesion molecules, cytokines and Enzymes as collective factors, mostly contributing to inflammatory response [6]. Chronic iron overload (CIO) can enhance proliferation and increase TNF-α. The latter can also induce nitric oxide (NO) production from the activated polmorphonuclear leukocytes in response to tissues inflammation [7,8]. NO is an inorganic reactive nitrogen species synthesized in liver by iNos located in hepatocytes, kupffer, endothelial cells [9]. It represents a hepatoprotective mechanism against

CIO toxicity [10]. Decreased level of Testosterone in male suffering from idiopathic hemochromatosis was reported before [11]. Androgenic deficiency as mentionned there not only produced by pituitary failure but also by testicular dysfunction due to hemosiderin deposition in testes. Oxidative stress may suppress also steroidogenesis. This is through substantial decrease in mRNA of steroid Acute regulatory protein (Star) as well as activities of testicular $\Delta^5$-3ß and 17-ß hydroxysteroid dehydrogenases [12]. Therefore present study aimed mainly to illustrate the effect of prolonged intake of iron enriched biscuits on oxidative stress, inflammation, testicular injuries and its degree in experimental rats. This may add great focus on such program having higher fortification degree with iron.

## 2. MATERIALS AND METHODS

A total of 30 wistar male albino rats (180 ± 20 g, 3 months old) were supplied by the Egyptian Organization for Biological Products and Vaccines (Helwan farm, Cairo). Rats were housed in stainless steel rodent cages and kept under environmentally controlled conditions and allowed one week for acclimatization at room temperature with a 12 h dark/light cycle before starting the experimental work. During such time animals received normal rat chow (El-Nasr Pharmaceuticals Chemicals, Egypt), and allowed free access of drinking water.

## 3. WORK DESIGN

The rats were then divided into 2 groups, first one received in addition to normal chow, iron enriched biscuits (0.3% ferrous sulphate w/w) [13] daily for ten weeks (Fe group). Second group received iron free biscuits and served as control. This protocol was approved by the animal care and use committee of the Biochemistry Department, Faculty of Pharmacy, Zagazig University. Both groups, approximately consumed equal quantities of biscuits, its main components were: wheat flour, eggs, Butter, cane sugar and vanillin as flavor. Ten weeks after feeding, rats were fasted overnight and blood was collected via retro-orbital bleeding. Serum was prepared, and aliquots were stored at –20℃ for later determinations of: iron, ferritin, TNF-α, testosterone and nitric oxide (NO).

Rats were killed by dissection. Testicular tissues were removed, rinsed with cold normal saline, divided into parts and dried with filter paper. First part was quickly frozen in liquid nitrogen (–170℃) then stored at –20℃ for determination of biochemical parameters: malondialdehyde (MDA), Glutathione (GSH) Hydroxyproline (Hyp) and iron contents.

The other part was kept in 10% formalin-saline at 4ºC for 1 week, subsequently dehydrated with a series of ethanol solution from 75 to 100% before embedding in paraffin. Cross sections (5 µm thick) were stained with hematoxylin and eosin (H & E) for routine light microscopy assessments, Perl's Prussian blue stain to localize deposited iron and Mallory trichrome stain to illustrate collagen fibers.

## 4. ANALYTICAL PROCEDURES

I: Serum iron was determined Colorimetrically by using The kits supplied by spinreact, (S.A., Spain) [14], NO as nitrite [15]. TNF-α was measured by using ELISA kits supplied by Biosource Int. (California, USA) [16]. Serum testosterone and ferritin were measured by electrochemiluminscence immunoassay "ECLIA" using commercial kits (Roche Diagnostics, USA) and Roche Elecsys 2010 immunoassay analyzers [17,18].

II: Testicular tissues: MDA was determined spectrophotometrically by using thiobarbituric acid, TBA [19]. 0.5 G tissue was homogenized in 5 ml phosphate buffer (PH = 7.2), centrifuged at 2000 g for 10 min, supernatant fraction was used for MDA determination. GSH content was determined spectrophotometrically using Ellman's reagent according to modified method [20]. 0.1 G of tissues was homogenized in 1 ml phosphate buffer (PH = 8) at 4℃. 0.5 ml of homogenate was mixed with 0.5 ml of 10 percent TCA in 5 mM EDTA sodium, mixed well and centrifuged at 2000 g for 5 min. Supernatant was used for determination of reduced GSH. Iron content was determined by flame atomic absorption Spectrophotometer [21]. About 0.1 G tissues was digested in 2ml conc Nitric acid and 2 ml perchloric acid at room temperature for 24 hours, the samples were filtered, diluted and absorption was measured at 248 nm. Hyp was determined spectrophotometrically by Ehrlich reagent [22]. 0.1 G tissues was pulverized ground, 500 ul of 6 N Hcl were added and incubated overnight at 120 c. 5 ul of acid hydrolysate were mixed with 5 ul of citrate acetate buffer and 100 ul chloramines T in ELISA plate and incubated for 20 minutes at room temperature before addition of Ehrlich solution.

## 5. STATISTICAL ANALYSIS

Analysis was carried out by using SPSS program for windows version 10 (SPSS, Chicago, USA). All results are presented as "Mean ± SD", student "t" test was used for the comparison between groups. Statistical significance was defined at $P < 0.05$.

## 6. RESULTS

Rats received biscuits enriched with iron (BEI) demonstrated significant increase in serum iron, ferritin, TNF-α, NO and decreased testosterone level ($p < 0.05$) (**Table 1**). Testicular contents of MDA, Hyp, and total iron demonstrated significant increase while GSH showed significant decrease ($p < 0.05$) (**Table 2**).

Serum iron showed positive correlation with ferritin r = 0.8, TNF-α r = 0.57 and negative one with testosterone r = –0.72 ($p < 0.05$). Testicular iron content was positively correlated with MDA r = 0.948, Hyp r = 0.978 and negatively with GSH r = –0.861 ($p < 0.05$).

The histopathological changes associated with Fe are illustrated in (**Figures 1-6**).

**Table 1.** Serum parameters of rats received iron fortified biscuits.

| Item | Control | BEI Group |
|---|---|---|
| Iron (mg/dl) | 264.6 ± 44.5 | 518.45 ± 90.2# |
| Ferritin (ng/ml) | 4.83 ± 0.76 | 10.58 ± 90.2# |
| Testosterone (ng/dl) | 123.63 ± 24.91 | 43.92 ± 9.3 # |
| NO (µmol/l) | 34.23 ± 4.69 | 66.37 ± 6.38 # |
| TNF-α (pg/ml) | 130.98 ± 7 | 213.6 ± 4.33# |

# Significantly different from control at $P < 0.05$.

**Table 2.** Testicular parameters of rats received iron fortified Biscuits.

| Item | Control | BEI |
|---|---|---|
| MDA (nmol/g tissue) | 669.17 ± 49.64 | 1439 ± 70.58# |
| GSH (nmol/g protein) | 38.09 ± 1.93 | 18.63 ± 2.51# |
| Hyp (µg/g tissue) | 130.98 ± 7 | 213.6 ± 4.33# |
| Iron (µg/gtissue) | 122.67 ± 7.84 | 265.17 ± 11# |

# Significantly different from control at $P < 0.05$.

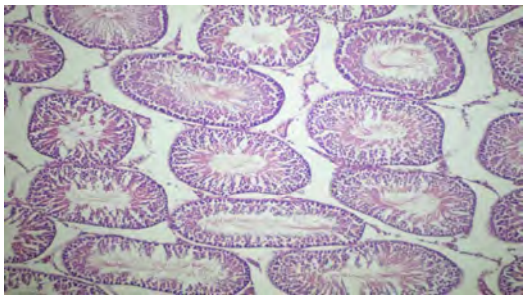


**Figure 1.** A photomicrograph of adult male albino rat testis (control group) showing seminiferous tubules (arrow) separated by interstitial tissue (double arrow) (H & E × 100).

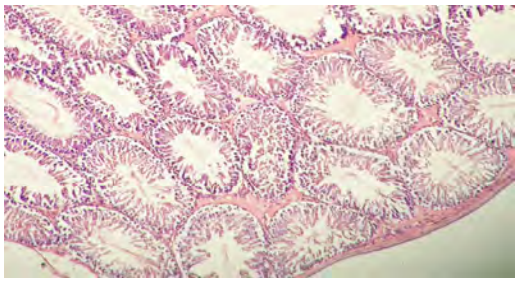


**Figure 2.** A photomicrograph of adult male albino rat testis of iron overload group showing adhesion of seminiferous tubules. Some tubules revealed disorganized germinal epithelium (arrow). Extensive area of exudates can be seen (double arrow) (H & E × 100).

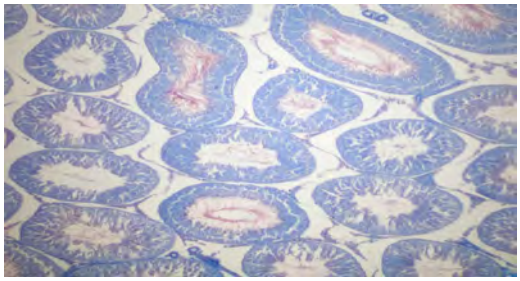


**Figure 3.** A photomicrograph of adult male albino rat testis (control group) showing distinctive boundary formed of collagen fibers (Mallory trichrome × 100).

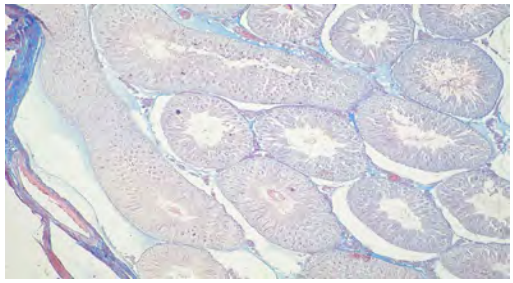


**Figure 4.** A photomicrograph of adult male albino rat testis (iron overload group) showing multiple collagen fibers (Mallory trichrome × 100).

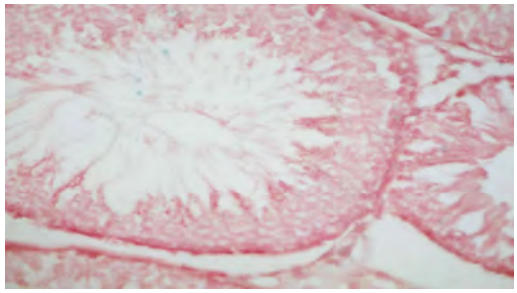


**Figure 5.** A photomicrograph of adult male albino rat testis (control group) showing negative Perl's Prussian blue stain (Perl's Prussian blue × 400).

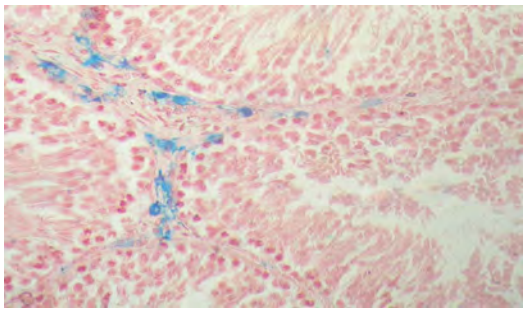


**Figure 6.** A photomicrograph of adult male albino rat testis (iron overload group) showing positive Perl's prussian blue stain (arrow) in interstitial tissue (Perl's Prussian blue × 400).

## 7. DISCUSSION

As mentioned above prolonged intake of iron enriched diet induced significant increase in serum iron and ferritin levels, in agreement with reported studies [23,24]. Previous report suggested that iron overload in rats can specifically activate target genes in the liver (*i.e.*, L ferritin and procollagen) [25]. Similar study supported this suggestion, illustrated an increased expression of the heavy-chain isoform of ferritin mRNA in liver of iron overloaded rats [26]. Others suggested that the elevated non-transferrin bound iron (NTB iron); observed in iron-overloaded diseases may reflect two factors. First one,by increasing saturation degrees of transferrin it will reduce the number of plasma binding sites for NTB iron .This may increase the total plasma iron fraction in the NTB iron pool. The second factor may be liver damage [27].

Testicular iron content demonstrated significant increase (**Table 2**), in agreement with previous studies [28,29] which considerd testicular tissues as secondary target for accumulated iron. This observation also implicates a regulation of iron transport which greatly differs from liver. It considers also that testis produce their own transferrin, and participates in iron shuttle system which fulfill the requirements of this metal for spermatogenesis [30].

Pro-oxidant effects of iron in testicular tissue are reasonable factor which mediates MDA increase and GSH decrease [31]. Other studies demonstrated GSH depletion, reduced activities of antioxidant enzymes like glutathione peroxidase, catalase, and superoxide dismutase [32]. In vitro study reported also similar findings [33].

Testicular content of Hyp demonstrated also significant increase in consistent with. Previous study [34] which may be due to increased activity of prolyl hydroxylase [35]. It was demonstrated before that iron overload induced moderate fibrosis in testes interstitium, interstitial macrophages, fibroblast and Leydig cells [36] meanwhile present study configurated similar findings (**Figure 4**).

Last years many studies has been oriented to illustrate iron-induced oxidative stress, its effect on activation of the transcription factor nuclear factor-kappa B (NF-kB) [37]. NF-kB exists in an inactive state in the cytosol, usually as two subunits, p50and p65, complexed with inhibitory proteins (IkB). Degradation of inhibitory proteins can translocate NF-kB to the nucleus where it alters transcription of a number of genes involved in cell growth, differentiation, inflammation, and immune function. Such study suggested that NF-kB activation can be modulated by the redox status of the cell [38]. ROS can increase NF-kB activation while, antioxidant compounds including iron chelators can inhibit it [39].

NF-kB activation also plays a key role in the inflammatory process and upregulates the transcription of adhesion molecules, cytokines and enzymes involved in the inflammatory responses [40]. Present study demonstrated significant increase in TNF-α of iron group and, in agreement with recent study [13].

TNF-α may also activate NO production from the activated polymorphnuclear leukocytes in response to tissue inflammation [6]. The presence of nitrogen reactive species can explain the iron sequestration pattern which characterizes macrophages under inflammatory conditions.It may refer also to a complex relationship between iron and NO so that iron overload can also increase iron uptake by the liver Kupffer cells while No increase is mostly attributed to activated nitric oxide synthase [41].

NO expression is controlled also by the redox-sensitive transcription factor NF-KB [10], acting in iron overload state as hepatoprotective agent against iron toxicity.

This may be associated with increased NF-kB DNA binding to the iNOS gene promoter and higher expression and activity of iNOS mRNA [42].

Decreased serum testosterone in Fe group may be attributed to haemosiderin deposition in the pituitary gland [43] or cytotoxic effects of iron on the gonadotrophs [44].

Reported histopathological study indicated that excessive iron deposition caused degranulation of the adenohypophysiocytes and decreased hormone storage within such cells [45], additionally iron deposition in the anterior lobe of the pituitary gland can decrease its response to gonadotropin-releasing hormone [12]. Involvement of oxidative stress in the suppression of steroidogenesis was also reported before. This may be due to substantial reduction in the mRNA of steroid acute regulatory protein as well as activities of testicular $\Delta^5$-3β and 17-β hydroxysteroid dehydrogenases (HSD) through strong affinity of divalent heavy metal to the thiol groups of these proteins and enzymes [46].

Other reported studies indicated that TNF-α induced

an early and transient increase in serum luteinizing hormone (LH) , followed by a transient decrease in serum testosterone levels, while FSH demonstrated no change [47].

## 8. CONCLUSIONS

Taken together, the present study indicates that iron overload causes impairment in testicular activity and affects the androgenicity of male rats .This may be a reflection of iron deposition in testicular tissues, a matter which deserve great focus and attentions regarding such dietary program which have higher fortification degree with iron.

## 9. ACKNOWLEDGEMENTS

We acknowledge the financial support (Research grant) for the study presented from Biochemistry Department, Faculty of Pharmacy-Zagazig University.

## REFERENCES

[1] Boldt, D.H. (1999) New perspectives on iron: An introduction. *American Journal of the Medical Sciences*, **318(4)**, 207-212.

[2] Houglum, K., Filip, M, Witztum, J.L. and Chojkier, M. (1990) Malondialdehyde and 4-hydroxynonenal protein adducts in plasma and liver of rats with iron overload. *Journal of Clinical Investigation*, **86(6)**, 1991-1998.

[3] Toyokunis, J. (2002) Iron and carcinogenesis: From fenton reaction to target genes. *Redox Report*, **7(4)**, 189-197.

[4] Swanson, C.A. (2003) Iron intake and regulation: Implications for iron deficiency and iron overload. *Alcohol*, **30(2)**, 99-102.

[5] Papanikolaou, G. and Pantapoulos, K. (2005) Iron metabolism and toxicity. *Toxicology and Applied Pharmacology*, **202(2)**, 199-211.

[6] Richard, A., Johnson, A.C. and Hanson, S. (2005) Parenteral iron compounds sensitize mice to injury-initiated TNF-α mRNA production and TNF-α release. *American Journal of Physiology. Renal Physiology*, **57(2)**, F290-F297.

[7] Brown, K.E., Meleah, M., Keimberly, A. and Weydert, J. (2006) Chronic iron overload stimulates hepatocyte proliferation and cyclin D1 expression in rodent liver. *Translational Research*, **148(2)**, 55-62.

[8] Dunger, A., Cunningham, J.M., Delaney, C.A. and Lowe, E. (1990) Tumor necrosis factor-α and Interferone-y inibit insulin secretion and cause DNA damage in unweanled rat islets. *Diabetes*, **45(2)**, 183-189.

[9] Alderton, W., Cooper, C. and Knowles, R. (2001) Nitric oxide synthases: Structure, function and inhibition. *Biochemical Journal*, **357(Pt 3)**, 593-615.

[10] Cornejero, P., Fernandez, V., Vial, M. and Videla, L.A. (2007) Hepatoprotective role of NO in an experimental model of chronic iron overload. *Nitric Oxide*, **16(1)**, 143-149.

[11] Vogt, H.J., Weidenbach, T., Marqart, K.H. and Vogel, G. (1987) Idiopathic hemochromatosis in a 45 year old infertile man. *Androlgia*, **19(5)**, 532-538.

[12] Gupta, R.S., Gupta, E.S., Dhakal, B.K. and Thakur, A.R. (2004) Vitamin C and Vitamin E protect the rat testes from cadmium-induced reactive oxygen species. *Mol cells*, **17(1)**, 132-139.

[13] Elmegeed, G.A., Ahmed, H. and Hussein, J.S. (2005) Novel synthesized aminsteroidal hetrocycles intervention for inhibiting iron induced oxidative stress. *European Journal of Medicinal Chemistry*, **40(12)**, 1283-1294.

[14] Burits, C.A. and Ashwood, E.R. (1999) Methods for the determination of serum iron, iron binding capacity, and transferrin saturation. In: Burtis, C.A., Edward, R., Ashwood, M.D. and Tietz, N.W., Eds., *Tietz Textbook of Clinical Chemistry*, 3rd Edition, AACC, 1701-1703.

[15] Moshage, H., Kok, B., Huizenga, J.R., Jansen, P.L.M. (1995) Nitrite and nitrate determination in plasma: A critical evaluation. *Clinical Chemistry*, **41(6 Pt 1)**, 892-896.

[16] Chen, W., Jin, W., Cook, M., Weiner, H.L. and Wahl, S.M. (1998) Oral delivery of group a streptococcal cell walls augments circulating TGF-beta and suppresses streptococcal cell wall arthritis. *The Journal of Immunology*, **161(11)**, 6297-6304.

[17] Wheeler, M.J. (1995) The determination of bio-available testosterone. *Annals of Clinical Biochemistry*, **32(Pt 4)**, 345-357.

[18] Lotz, J., Hafner, G. and Prellwitz, W. (1997) Reference study for ferritin assays. *Kurzmitteilung Clinical Laboratory*, **43(11)**, 993-994.

[19] Buege, J. and Aust, S. (1978) Microsomal lipid peroxidation. *Methods Enzymology*, **52**, 302-306.

[20] Ahmed, A.E., Gamal, I.H., Loh, J. and Abdel-Rahman, S.Z. (1991) Studies on mechanism of haloacetonitrile induced gasyrointestinal toxicity, interaction of dibromoacetonitrile with glutathione as glutath ione-s-transferase in rats. *Journal of Biochemical Toxicology*, **6(2)**, 115-121.

[21] Basset, M.L., Halliday, J.W. and Powell, L.W. (1986) Value of hepatic measurement in early hemochromatosis and determination of the critical iron level associated with fibrosis. *Hepatology*, **6(1)**, 24-29.

[22] Fujita, M., Shannon, J.M., Morikawa, O., Gauldie, J., Hara, N. and Mason, R.J. (2003) Overexpression of tumor necrosis factor-α diminishes pulmonary fibrosis induced by bleomycin or transforming growth Factor-β. *American Journal Respiratory Cell and Molecular Biology*, **29(6)**, 669-676.

[23] Zhao, Y., Li, H., Gao, Z. and Xu, H. (2005) Effects of dietary baicalin supplementation on iron overload induced mouse liver oxidative injury. *European Journal of Pharmacology*, **509(2-3)**, 195-200.

[24] Silvana, M.L., Ribeiro, T., Sliva, M.E., Chianca, D.A., Paula, H.D., Cardoso, L.M., Colombari, E. and Pedrosa, L. (2003) Iron overload in hypercholesterolemic rats affects iron homeostasis and serum lipids but not blood pressure. *Journal of Nutrition*, **133(1)**, 15-20.

[25] Pietrangelo, A., Rocchi, E., Schiaffonati, L., Ventura, E. and Cairo, G. (1990) Liver gene expression during chronic dietary iron overload in rats. *Hepatology*, **11(5)**, 798-804.

[26] Valerio, L.G.Jr. and Petersen, D.R. (2000) Characterization of hepatic iron overload following dietary administration of dicyclopentadienyl iron (ferrocene) to mice: Cellular, biochemical, and molecular aspects. *Experimental and Molecular Pathology*, **68(1)**, 1-12.

[27] Teresa, L., Brissot, P., Ma, W.-L. and Weisiger, R.A. (1986) Characterization of non-transferrin-bound iron clearance by rat liver. *The Journal of Biological Chemistry*, **261(23)**, 10909-10914.

[28] Lucesoli, F. and Fraga, C.G. (1999) Oxidative stress in testes of rats subjected to chronic iron intoxication and α-tocopherol supplementation. *Toxicology*, **132(2-3)**, 179-186.

[29] Galleano, M. and Puntarulo, S. (1997) Dietary α-tocopherol supplementatiom on antioxidant defenses after in vivo iron overload in rats. *Toxicology*, **124(1)**, 73-81

[30] Sylvester, S.R. and Griswold, M.D. (1993) Molecular biology of iron transport in the testis. In: De Krester, D., Ed., *Molecular Biology of the Male Reproductive System*, Academic Press, San Diego, 311-323.

[31] Harandi, A.A., Allameh, A. and O'Brien, P.J. (2005) In vivo effects of iron overload on toxicological parameters in isolated hepatocytes obtained from adult rats. *FEBS Journal*, **272(s1)**, 1-650.

[32] Jagetia, G.C., Reddy, T.K., Venkatesha, V.A. and Kedlaya, R. (2004) Influence of naringin on ferric iron induced oxidative damage in vitro. *Clinica Chimica Acta*, **347(1-2)**, 189-197.

[33] Alidoost, F., Gharagozloo, M., Bagherpour, B., Jafarian, A., Sajjadi, S.E., Hourfar, H. and Moayedi, B. (2006) Effects of silymarin on the proliferation and glutathione levels of peripheral blood mononuclear cells from beta-thalassemia major patients. *International Immunopharmacology*, **6(8)**, 1305-1310.

[34] Zhang, Y., Li, H., Zhao, Y. and Gao, Z. (2006) Dietary supplementation of baicalin and quercetin attenuates iron overload induced mouse liver injury. *European Journal of Pharmacology*, **535(1-3)**, 263-269.

[35] Poli, G. and Parola, M. (1997) Oxidative damage and fibrogenesis. *Free Radical Biology and Medicine*, **22(1-2)**, 287-305.

[36] Lucesoli, F., Caligiuri, M., Roberti, M.F., Perazzo, J.C. and Fraga, C.G. (1999) Dose-dependent increase of oxidative damage in the testes of rats subjected to acute iron overload. *Archives of Biochemistry and Biophysics*, **372(1)**, 37-43.

[37] Tsukamoto, H., Lin, M., Ohata, M., Giulivi, C., French, S.W. and Brittenham G. (1999) Iron primes hepatic macrophages for NF-kB activation in alcoholic liver injury. *American Journal of Physiology*, **277(6 Pt 1)**, G1240-G1250.

[38] Flohe´, L., Brigelius-Flohe, R., Saliou, C., Traber, M.G. and Packer, L. (1997) Redox regulation of NF-kB activation. *Free Radical Biology and Medicine*, **22(6)**, 1115-1126.

[39] Li, S., Li, X. and Rozanski, G.J. (2003) Regulation of glutathione in cardiac myocytes. *Journal of Molecular and Cellular Cardiology*, **35(9)**, 1145-1152.

[40] Dunger, A., Cunningham, J.M., Delaney, C.A., Lowe, J.E., Green, M.H.L., Bone, A.J. and Green, I.C. (1996) Tumor necrosis factor-α and interferone-γ inhibit insulin secretion and cause DNA damage in unweanted-rat islets. *Diabetes*, **45(2)**, 183-189.

[41] Galleano, M., Simontacchi, M. and Puntarulo, S. (2004) Nitric oxide and iron: Effect of iron overload on nitric oxide production in endotoxemia. *Molecular Aspects of Medicine*, **25(1-2)**, 141-154.

[42] Kleinert, H., Pautz, A., Linker, K. and Schwarz, P. (2004) Regulation of the expression of inducible nitric oxide synthase. *European Journal of Pharmacology*, **500(1-3)**, 255-266.

[43] Seracchioli, R., Porcu, F. and Colombi, C. (1994) Transfusion-dependent homozygous β-thalassemia major successful twin pregnancy following in-vitro fertilization and tubal embryo transfer. *Human Reproduction*, **9(10)**, 1964-1965.

[44] Sparacia, G., Iaia, A., Banco, A., D'Angelo, P. and Lagalla, R. (2000) Transfusional hemochromatosis: Quantitative relation of MR imaging pituitary signal intensity reduction to hypogonadotropic Hypogonadism. *Radiology*, **215(3)**, 818-823.

[45] Bergeron, C. and Kovacs, K. (1978) Pituitary siderosis: A histologic, immunocytologic and ultrastructural study. *American Journal of Pathology*, **93(2)**, 295-310.

[46] Wang, C., Tso, S. and Todd, D. (1989) Hypogonadotropic hypogonadism in severe beta-thalassemia: Effect of chelation and pulsatile gonadotropin-releasing hormone therapy. *The Journal of Clinical Endocrinology and Metabolism*, **68(3)**, 511-516.

[47] van der Poll, T., Romijn, J.A., Endert, E. and Sauerwein, H.P. (1993) Effects of tumor necrosis factor on the hypothalamic-pituitary-testicular axis in healthy men. *Metabolism*, **42(3)**, 303-307.

**Vol.2, No.6, 557-562 (2010)**
doi:10.4236/ns.2010.26070

**Natural Science**

# Poly (ethylene terephthalate) synthesis with catalysts derived from chrysotile asbestos

**Shigeki Habaue[1*], Yusuke Takahashi[2], Yu Hosogoe[2], Hiroshi Yamashita[2], Meisetsu Kajiwara[3]**

[1]Faculty of Health and Nutrition, Shubun University, Ichinomiya, Japan; *Corresponding Author: habaue@shubun-ac.jp
[2]Department of Chemistry and Chemical Enginerring, Graduate School of Science and Engineering, Yamagata University, Yonezawa, Japan
[3]School of Dentistry, Aichi-Gakuin University, Nagoya, Japan

Received 1 March 2010; revised 30 April 2010; accepted 13 May 2010.

## ABSTRACT

**The chrysotile asbestos was converted to the forsterite-type compounds by calcination at 740 and 800℃ (F7-740 and F7-800), which were used as a catalyst for the polycondensation of bis-(hydroxyethyl) terephthalate affording poly (ethylene terephthalate). The obtained forsterite-type compounds did not show any catalytic activity. However, the products obtained by simply treating them with acetic acid significantly pro- moted the polymerization that produced a THF- insoluble polymer. It was found that the polymer prepared with the acetic acid-treated F7-740 at 160℃ for 2 h showed a 93% yield and the number average molecular weight of 6.4 × $10^3$. The observed catalytic activity was higher than that for the acetic acid-treated magnesium oxide, as well as the typical polycondensation catalysts, such as magnesium acetate and antimony oxide.**

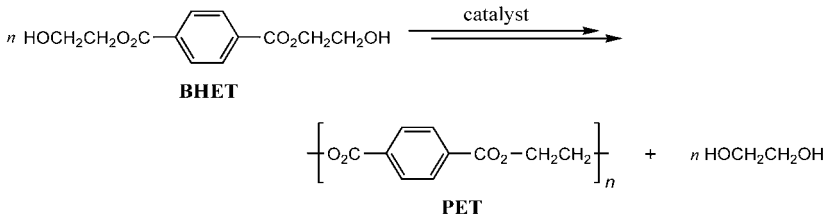
**Keywords:** Chrysotile Asbestos Forsterite; Polycondensation; Poly (Ethylene Terephthalate); Recycling

## 1. INTRODUCTION

Poly(ethylene terephthalate) (PET) is a versatile thermoplastic resin and extensively used for various products in the forms of fibers, films, etc. The PET is generally produced by the esterification of terephthalic acid with ethylene glycol followed by polycondensation, as well as by transesterification using dimethyl terephthalate with ethylene glycol or using bis(hydroxyethyl) terephthalate (BHET) as the starting material. The latter system with BHET is also important as one of the recycling processes of PET wastes, which is shown in **Scheme 1**.

During these polycondensation processes, antimony compounds, such as the oxide and acetate, are typically employed as the catalyst [1-4], and the replacing these heavy metal catalysts with others that are safe, economic, and highly catalytic has been desired [5-7]. For example, it was recently reported that hydrotalcite, $[Mg_6Al_2(OH)_{16}](CO_3)\cdot 4H_2O$, can be an efficient catalyst for the PET synthesis [8,9].

The chrysotile asbestos, represented by the approximate composition of $Mg_3Si_2O_4(OH)_4$, is comprised of a silica tetrahedral sheet [$SiO_4$] joined into a brucite layer of basic magnesium hydroxide [$Mg(OH)_2$], and the curvature of these layers affords a structure of tubular and cylindrical rolls with nanometer-order diameters [10]. The chrysotile asbestos has been widely applied in industry as a material having excellent physical and chemical properties, such as tensile strength, heat-resistance, durability, etc. However, its use is prohibited or strictly regulated at present, because of health hazards, that is, asbestosis and carcinogenesis of respiratory systems.




**Scheme 1.** Polycondensation of BHET producing PET.

The fibrous and needle-like structure of the asbestos is considered to be the main cause of these serious diseases. A vast amount of chrysotile asbestos has already been used, therefore, the development of potential disposal methods for these waste materials and the unused raw chrysotile is an essential and urgent subject. Especially, their transformation into nontoxic and valuable materials is of great significance.

Recently, we reported that chrysotile asbestos is effectively converted into polysiloxanes through selective acid-leaching and silylation [11,12]. This result offers a novel method for converting the hazardous chrysotile asbestos wastes into valuable polymer materials. The calcination of the chrysotile asbestos is another method affording a harmless material, which mainly consists of forsterite, $Mg_2SiO_4$, as shown in **Scheme 2** [13-15]. Accordingly, effective utilization of this material can also provide a certain way to solve this serious issue.

$$2\ Mg_3Si_2O_5(OH)_4 \longrightarrow 3\ Mg_2SiO_4 + SiO_2 + 4\ H_2O$$

**Scheme 2.** Calcination process of chrysotile asbestos.

In this study, the forsterite-type compounds prepared by the calcination of the chrysotile asbestos were evaluated as the polycondensation catalyst for BHET, and it was found that they show a significant catalytic activity by treating with carboxylic acids. Therefore, this novel catalyst system for the PET production from BHET can simultaneously contribute to the effective recycling both of the asbestos and PET wastes.

## 2. EXPERIMENTAL

### 2.1. Materials

The class 7 chrysotile (F7) [mined in Furano, Hokkaido, Japan (Nozawa Co.)] was used as the starting material, and the calcination at 800 or 740℃ for 1 h was conducted to produce a pale brownish, forsterite-type compound, F7-800 or F7-740 (caution!: the asbestos must be carefully treated, because the fibrils cause serious health hazards). The silica (L-$SiO_2$) with a fibrous structure originating from the chrysotile was prepared by the acid-leaching of the serpentine slag with sulfuric acid as previously reported [11,12]. During this process, almost all of the magnesium constituent was removed, that is, the brucite layer was effectively leached. Its chemical composition (%, fluorescent X-ray) was estimated to be 96.6, $SiO_2$; 2.4, $Al_2O_3$; 0.2, $Fe_2O_3$; 0.1, CaO, and it has a specific surface area of 178 $m^2/g$. The monomer, BHET (TCI), and the salts of magnesium and antimony, such as MgO, $Mg(OH)_2$, magnesium acetate [$Mg(OAc)_2$] (Kanto), $MgSO_4$ (Wako), and $Sb_2O_3$ (Nihon Seiko), were used as received.

### 2.2. Treatment of Forsterite with Acid

A mixture of the forsterite-type compound (0.20 g, F7-800 or F7-740) and acetic acid (7 equiv. based on Mg) was stirred for 24 h under a $N_2$ atmosphere at ambient temperature. After evaporation of the acid, the product was further vacuum-dried at 50℃ for 12 h. The obtained solid was used for the polycondensation reaction as the catalyst without further purification.

### 2.3. Polymerization

A mixture of BHET (0.50 g) and a catalyst (1 wt%) was reacted at 160℃ for 2 h under reduced pressure (< 0.5 mmHg). After cooling to room temperature, the reaction mixture was washed with methanol, then with tetrahydrofuran (THF). The insoluble fraction was isolated by centrifugation and dried *in vacuo*.

### 2.4. Measurements

The chemical compositions were determined by X-ray fluorescence using a Shimadzu EDX-800 spectrometer. The specific surface area was measured by the BET method. The powder X-ray diffraction (XRD) patterns were obtained using monochromatic CuKa radiation with a Rigaku RINT-2100-ultra diffractometer. The $^1H$ NMR spectra were measured by a Varian Unity Inova (500 MHz) spectrometer. The infrared (IR) spectra were recorded by a Horiba FT-720 spectrometer. The size exclusion chromatography (SEC) analysis was performed by a Jasco PU-2080-plus equipped with a Jasco UV-2075-plus UV detector with Shodex AC8025 and TSK-GEL columns connected in series (eluent = $CHCl_3$, flow rate = 1.0 mL/min). The polymers were dissolved in $CHCl_3$ by addition of a small amount of 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) and calibration was carried out with standard polystyrenes.

## 3. RESULTS AND DISCUSSION

### 3.1. Calcination of Chrysotile Asbestos

The XRD patterns of F7-800 and F7-740, prepared by the calcination of the chrysotile asbestos at 800 and 740℃ for 2 h under atmospheric pressure, are shown in **Figure 1**. The peaks based on the chrysotile completely disappeared during the calcination, and peaks due to the formation of forsterite were observed for both products. The compound calcined at the lower temperature, F7-740, showed a broader peak pattern than that of F7-800, suggesting that the calcination temperature affects the

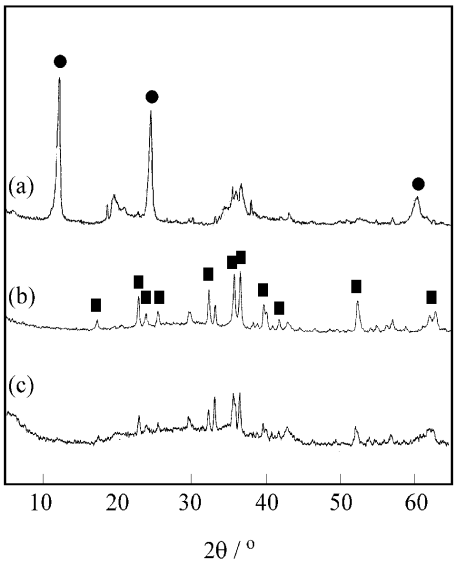



**Figure 1.** X-ray powder diffraction patterns of (a) chrysotile asbestos (serpentine slag), (b) F7-800, and (c) F7-740 (●: chrysotile, ■: forsterite).

crystallinity of the products. The product, F7-740, should have a lower crystallinity than F7-800.

The chemical composition, determined by a fluorescent X-ray analysis, and specific surface area of the starting material F7 and the calcined products, F7-800 and F7-740, are listed in **Table 1**. The amounts MgO, $SiO_2$, $Fe_2O_3$, $Al_2O_3$, and CaO in the products (F7-800 and F7-740) increased by the calcinations when compared to those of F7. This is due to the dehydration as shown in **Scheme 2**. The observed chemical compositions for the calcined products are quite similar to each other, while the BET surface area of F7-740, 25.1 $m^2/g$, was higher than that of F7-800. These results showed that the calcination temperature should affect the crystallinity. In other words, the product, F7-740, is rich in an amorphous phase.

## 3.2. Polycondensation of Bis(hydroxyethyl) Terephthalate

The polymerization of BHET at 160℃ under vacuum was carried out (**Table 2**). The polymerization without a catalyst did not proceed (entry 1). The typical magnesium and antimony salts, such as MgO, $Mg(OH)_2$, $Mg(OAc)_2$ and $Sb_2O_3$, except for $MgSO_4$, as the catalyst promoted the polymerization and produced the THF-insoluble polymers (entries 4-8). For example, $Mg(OAc)_2$ afforded a polymer in a good yield with the number average molecular weight ($M_n$) of $6.2 \times 10^3$. However, the forsterite-type compounds, F7-800 and F7-740, which mainly contain the magnesium constituent, did not afford any polymeric compounds (entries 2 and 3).

**Table 1.** Characterization of starting materials.

| Material | Chemical composition (%)[a] | | | | | Surface area ($m^2/g$)[b] |
|---|---|---|---|---|---|---|
| | MgO | $SiO_2$ | $Fe_2O_3$ | $Al_2O_3$ | CaO | |
| F7 | 50.5 | 29.7 | 3.7 | 1.0 | 0.4 | — |
| F7-800 | 58.9 | 32.9 | 4.1 | 1.3 | 0.5 | 18.2 |
| F7-740 | 57.9 | 32.6 | 4.0 | 1.3 | 0.5 | 25.1 |

[a]Determined by fluorescent X-ray analysis. [b]Determined by BET measurement.

**Table 2.** Polymerization of BHET with various catalysts at 160℃.

| Entry | Catalyst[a] | Yield (%)[b] | Activity (g-product·g-cat$^{-1}$·h$^{-1}$) | $M_n \times 10^{-3}$[c] | $M_w \times 10^{-3}$[c] |
|---|---|---|---|---|---|
| 1 | — | 0 | — | — | — |
| 2 | F7-800 | 0 | — | — | — |
| 3 | F7-740 | 0 | — | — | — |
| 4 | MgO[d] | 21 | 26.3 | 3.4 | 4.0 |
| 5 | $Mg(OH)_2$[e] | 55 | 45.8 | 4.9 | 5.7 |
| 6 | $Mg(OAc)_2$ | 87 | 43.5 | 6.2 | 7.2 |
| 7 | $MgSO_4$[f] | 0 | — | — | — |
| 8 | $Sb_2O_3$ | 22 | 11.0 | 4.9 | 5.4 |
| 9 | L-$SiO_2$ | 0 | — | — | — |

[a]BHET: 0.50 g, catalyst: 5 mg. [b]MeOH- and THF-insoluble part. [c]Determined by SEC (polystyrene standards). [d]MgO: 2 mg. [e]$Mg(OH)_2$: 3 mg. [f]$MgSO_4$: 6 mg.

F7-800 and F7-740 were then treated with acetic acid (AcOH) at room temperature as stated in the experimental section, and the obtained compounds were used as the polycondensation catalyst. During this process, the weight increases of 1.8 × $10^2$ and 3.3 × $10^2$ mg for 1.0 g of F7-800 and F7-740, respectively, were observed. The IR spectra of these compounds are shown in **Figure 2**, together with those of F7-740 and $Mg(OAc)_2$. The absorptions based on the magnesium acetate around 1100-1300 $cm^1$ were clearly observed for F7-800 and F7-740 when treated with AcOH. Accordingly, during this process, the formation of the magnesium salt of AcOH should take place, whereas the yields were poor, judging from the observed weight increase.

On the other hand, MgO was also treated with AcOH (12 equiv.) to give a reaction mixture, which showed a weight increase of 3.0 × $10^3$ mg for 1.0 g of MgO. The leached silica, L-$SiO_2$, prepared from the serpentine slag [11,12], was calcined at 740℃ for 1 h, and it was also treated with 5 equiv. of AcOH. However, no weight increase was observed. These results were quite different from those observed for F7-800 and F7-740, supporting the fact that the magnesium constituent in the forsterite-type compounds should react with AcOH.

The results of the polycondensation of BHET with the catalyst (1 wt%), prepared from the forsterite-type compounds by treating with various organic carboxylic acids, are summarized in **Table 3**. The MgO derivative, obtained by treating with AcOH, produced a polymer in 84% yield and showed an catalytic activity of 42.0 g-product/g-cata- lyst·h (**Table 3**, entry 1), whose values were almost comparable to those of the polymerization using $Mg(OAc)_2$ (**Table 2**, entry 6). The acid-treated L-$SiO_2$, as well as the original L-$SiO_2$, produced no THF-insoluble fraction during the polymerization (**Table 3**, entry 11 and **Table 2**, entry 9).

In marked contrast, the F7-800 and F7-740 derivatives, treated with AcOH, resulted in good yields with the $M_n$ value of approximately 6.3 × $10^3$ (**Table 3**, entries 2 and 3), although F7-800 and F7-740 without acid-treatment showed no catalyst activity (**Table 2**, entries 2 and 3). Especially, the AcOH-treated F7-740 showed the higher

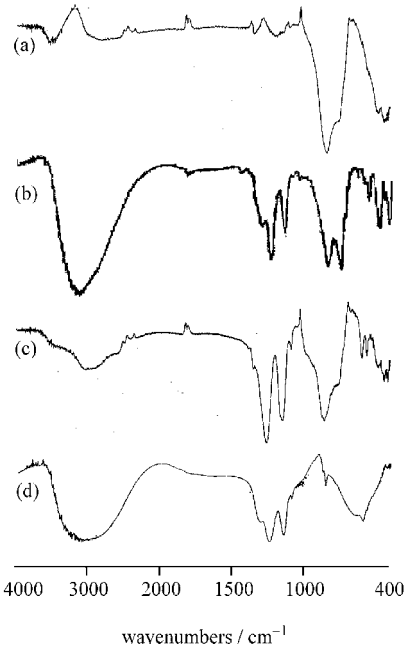



**Figure 2.** IR spectra of (a) F7-740, (b) F7-800 after treating with AcOH, (c) F7-740 after treating with AcOH at room temperature, and (d) $Mg(OAc)_2$ (KBr).

**Table 3.** Polymerization of BHET with various acid-treated materials at 160℃[a].

| Entry | Starting material | Acid[b] | Yield (%)[c] | Activity (g-product·g-cat$^{-1}$·h$^{-1}$) | $M_n$ × $10^{-3}$[d] | $M_w$ × $10^{-3}$[d] |
|---|---|---|---|---|---|---|
| 1 | MgO | AcOH | 84 | 42.0 | 5.3 | 5.9 |
| 2 | F7-800 | AcOH | 71 | 35.5 | 6.3 | 7.5 |
| 3 | F7-740 | AcOH | 93 | 46.5 | 6.4 | 7.6 |
| 4[e] | F7-740 | AcOH | 90 | 45.0 | 13.1 | 20 |
| 5[f] | F7-740 | AcOH | 68 | 56.7 | 5.1 | 5.6 |
| 6[f] | F7-740 | AcOH[g] | 77 | 64.2 | 5.0 | 5.5 |
| 7[f] | F7-740 | AcOH[h] | 80 | 66.7 | 5.2 | 5.8 |
| 8 | F7-740 | PA[h,i] | 46 | 23.0 | 5.7 | 6.9 |
| 9 | F7-740 | TFA[j] | 62 | 31.0 | 5.2 | 5.9 |
| 10 | F7-740 | TFA[g,j] | 48 | 21.5 | 4.7 | 5.3 |
| 11 | L-$SiO_2$[k] | AcOH | 0 | — | — | — |

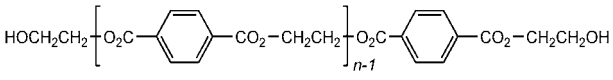
[a]BHET: 0.50 g, catalyst: 5 mg. [b]About 7 equiv. of acid was used. [c]MeOH- and THF-insoluble part. [d]Determined by SEC (polystyrene standards). [e]Reaction temp. = 220℃. [f]Catalyst: 3 mg. [g]F7-740 was treated with AcOH at 60℃. [h]F7-740 was treated with AcOH at 100℃. [i]Propionic acid. [j]Trifluoroacetic acid. [k]L-$SiO_2$ calcinated at 740℃ was used.

yield of 93% and a catalytic activity of 46.5 g-product/g-catalyst·h than those of the acid-treated MgO derivative, as well as those of $Mg(OAc)_2$, although the weight increase of F7-740 during the AcOH-treatment step was much lower than that of MgO. These results suggest that the activity of the catalyst sites in the F7-740, generated during the acid-treatment process, should be significantly high when compared to that of the acid-treated MgO and $Mg(OAc)_2$. Therefore, in addition to the formation of the magnesium salt of AcOH in F7-740, some structure of the forsterite, prepared from the chlysotile asbestos, should play an important role in the polycondensation of BHET.

The polymerization of BHET with the AcOH-treated F7-740 was performed at 220℃ (**Table 3**, entry 4). Although the observed polymer yield and catalytic activity were similar to those of the polymerization at 160℃, the obtained polymer showed an $M_n$ value of $1.3 \times 10^4$, which was estimated by SEC. Therefore, the polymerization at higher temperatures could be effective for this catalyst system.

The $^1$H NMR spectrum of the THF-insoluble part obtained using the AcOH-treated F7-740 (**Table 3**, entry 4) was depicted in **Figure 3**. The polymer was soluble in chloroform containing a small amount of HFIP, and the peaks were assigned as shown in the figure. The small peaks around 4.0 and 4.5 ppm were based on the methylene protons of the terminal hydroxyethoxy groups (**Scheme 3**). Accordingly, the $M_n$ value calculated from the integral ratio was $2.5 \times 10^3$, which is much lower than that estimated by SEC, probably due to the rigid aromatic main chain of the polymer and some aggregation.




**Scheme 3.** Structure of the obtained PET.

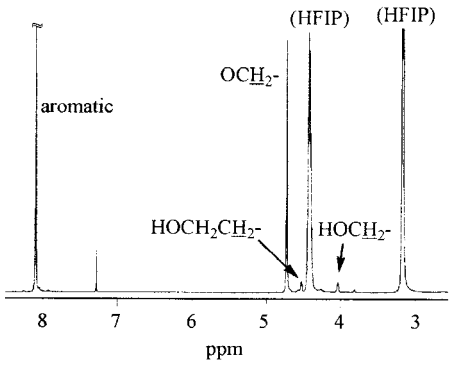



**Figure 3.** $^1$H NMR Spectrum of the obtained polymer (**Table 3**, entry 4) (in $CDCl_3$ containing HFIP).

The forsterite-type compound, F7-740, was treated with heated AcOH for 24 h. During this process, the weight increases of $6.1 \times 10^2$ and $9.3 \times 10^2$ mg/g-F7-740 for the reaction at 60 and 100℃, respectively, were observed. Therefore, the reaction between the magnesium in the compound and AcOH proceeded more effectively than the reaction at room temperature. The obtained materials were employed as the catalyst (0.6 wt%) for the polymerization of BHET at 160℃ (**Table 3**, entries 5-7). Under these reaction conditions, the acid-treated F7-740 at room temperature showed a catalytic activity of 56.7 g-product/g-catalyst·h, whereas the F7-740 derivatives, treated with AcOH at higher temperatures, showed significantly higher catalytic activities. For example, the highest activity of 66.7 g-product/g-catalyst·h was observed for the polymerization using the F7-740 derivative prepared at 100℃.

F7-740 was further ground and treated with AcOH at 100℃ for 24 h. The weight increase was $1.1 \times 10^3$ mg/g-F7-740, which is higher than that observed for the reaction of F7-740 without grinding. The polycondensation of BHET at 160℃ with this catalyst (0.6 wt%) was conducted, and the THF-insoluble fraction was obtained in a 72% yield with the catalytic activity of 60.0 g-product/g-catalyst·h and the $M_n$ of $5.4 \times 10^3$ as evaluated by SEC. This result was almost comparable to that of the polymerization using the acid-treated F7-740 without grinding.

Finally, the effect of the carboxylic acid on the catalytic activity during the BHET polymerization at 160℃ (catalyst: 1 wt%) was examined (**Table 3**, entries 8-10). The forsterite-type material, F7-740, was treated with propionic acid (PA) at 100℃. After evaporation of the unreacted acid, the weight increase was $9.4 \times 10^2$ mg/g-F7-740, of which the molar yield is lower than that observed for the reaction with AcOH at 100℃. The acidity of the carboxylic acid should have an effect on the reaction. The polymerization with this catalyst resulted in a much lower yield and catalytic activity than those for the polymerization using the AcOH-treated F7-740 catalyst. F7-740 was also treated with trifluoracetic acid (TFA) at room temperature or 60℃ affording a reaction mixture with the weight increase of $9.2 \times 10^2$ or $2.6 \times 10^3$ mg/g-F7-740, respectively. The latter catalyst preparation reaction at 60℃ significantly proceeded, however, it showed a low catalytic activity of 21.5 g-product/g-catalyst·h for the polymerization. Therefore, AcOH is currently the most effective carboxylic acid, that is, the magnesium salt of AcOH generated in the forsterite-type compound, F7-740, effectively promotes the polycondensation of BHET, although a detail structure of the active site is not clear at present.

## 4. CONCLUSIONS

The chrysotile asbestos was calcined affording the forsterite-type compounds. Although the polycondensation of BHET with them did not proceed, the compounds, just treated with acetic acid, showed a significant catalytic activity, which was higher than that for the polymerization by the MgO derivative treated with acetic acid, as well as typical polycondensation catalysts, such as $Mg(OAc)_2$ and $Sb_2O_3$. Some phases or amorphologies having the composition of the forsterite derived from the chrysotile asbestos by the calcination, in addition to the magnesium salt generated by acid-treatment, should be important for the catalytic activity during the polymerization. Accordingly, this novel catalyst system could contribute to both the disposal of chrysotile asbestos waste and the recycling of PET waste.

## REFERENCES

[1] El-Toufaili, F.-A., Feix, G. and Reichert, K.-H. (2006) Mechanistic investigations of antimony-catalyzed polycondensation in the synthesis of poly(ethylene terephthalate). *Journal of Polymer Science Part A*: *Polymer Chemistry*, **44(3)**, 1049-1059.

[2] Lee, S.W., Ree, M., Park, C.E., Jung, Y.K., Park, C.-S., Jin, Y.S. and Bae, D.C. (1999) Synthesis and non-isothermal crystallization behaviors of poly(ethylene isophthalate-co-terephthalate)s. *Polymer*, **40(25)**, 7137-7146.

[3] Zhang, Y. and Gu, L. (2000) Study of non-isothermal crystallization kinetics and sequence distribution in poly (ethyele terephthalate-co-isophthalate). *European Polymer Jounal*, **36(5)**, 759-765.

[4] Greener, J., Gillmor, J.R. and Daly, R.C. (1993) Melt rheology of a class of polyester ionomers. *Macromolecules*, **26(24)**, 6416-6424.

[5] MacDonald, W.A. (2002) New advances in poly(ethylene terephthalate) polymerization and degradation. *Polymer International*, **51(10)**, 923-930.

[6] Gorzawski, H. and Hoelderich, W.F. (1999) Transesterification of methyl benzoate and dimethyl terephthalate with ethylene glycol over superbases. *Applied Catalysis A*: *General*, **179(1-2)**, 131-137.

[7] Meyer, U. and Hoelderich, W.F. (1999) Transesterification of methyl benzoate and dimethyl terephthalate with ethylene glycol over basic zeolites. *Applied Catalysis A*: *General*, **178(2)**, 159-166.

[8] El-Toufaili, F.-A., Ahmadniana, F., Dinse, A., Feix, G. and Reichert, K.-H. (2006) Studies on hydrotalcite-catalyzed synthesis of poly(ethylene terephthalate). *Macromolecular Materials and Engineering*, **291(Pt 4)**, 1136-1143.

[9] Serio, M.D., Tesser, R., Ferrara, A. and Santacesaria, E. (2004) Heterogeneous basic catalysts for the transesterification and the polycondensation reactions in PET production from DMT. *Journal of Molecular Catalysis A*: *Chemical*, **212(1-2)**, 251-257.

[10] Falini, Foresti, G.E., Lesci, G. and Roveri, N. (2002) Structural and morphological characterization of synthetic chrysotile single crystals. *Chemical Communications*, **14**, 1512-1513.

[11] Habaue, S., Hirasa, T., Akagi, Y., Yamashita, K. and Kajiwara, M. (2006) Synthesis and property of silicone polymer from chrysotile asbestos by acid-leaching and silylation. *Journal of Inorganic and Organometallic Polymers and Materials*, **16(2)**, 155-160.

[12] Habaue, S., Sato, K., Yamashita, K., Shimamura, T., Kaito, M., Masuda, T. and Kajiwara, M. (2008) Polysiloxanes derived from chrysotile asbestos via acid-leaching and silylation processes, *Journal of Applied Polymer Science*, **110(5)**, 2891-2897.

[13] Gruner, J.W. (1948) Progress in silicate structures. *The American Mineralogist*, **33(9-10)**, 679-691.

[14] Brindley, G.W. and Zussman, J. (1957) A structural study of the thermal transformation of serpentine minerals to forsterite. *The American Mineralogist*, **42(8)**, 461-474.

[15] Arai, Y. and Nagai, S. (1963) Chemical Utilization of Serpentine. *Chemistry and Chemical Industry*, **16(1)**, 59-68.

**Vol.2, No.6, 563-570 (2010)**
doi:10.4236/ns.2010.26071

**Natural Science**

# Binding of naturally occurring hydroxycinnamic acids to bovine serum albumin

**Lucie Trnková[1,2], Iva Boušová[1*], Vladimír Kubíček[1], Jaroslav Dršata[1,2]**

[1]Charles University in Prague, Faculty of Pharmacy, Hradec Králové, Czech Republic; *Corresponding Author: Iva.Bousova@faf.cuni.cz
[2]University of Hradec Králové, Faculty of Education, Hradec Králové, Czech Republic

Received 22 February 2010; revised 12 April 2010; accepted 13 May 2010.

## ABSTRACT

**Hydroxycinnamic acids (HCAs) possess numerous biological effects including antioxidant, antiallergic, antimicrobial, and immunomodulatory activities and due to these properties are widely used in folk medicine. Nevertheless, they can interact with protein molecules and cause some structural and functional changes. The possibility of HCAs binding to bovine serum albumin (BSA) under physiological conditions was investigated by the UV-VIS absorption spectroscopy and fluorescence quenching method. Apart from rosmarinic acid, all tested HCAs quenched tryptophan fluorescence of BSA in the studied range of concentrations (0-20 μM) mainly by static quenching mechanism (formation of non-fluorescent HCA-BSA complexes). The binding constants, number of binding sites and free energy changes were determined. The binding affinities of HCAs were ranked in the order: chlorogenic acid > sinapic acid ≥ caffeic acid > ferulic acid > *o*-coumaric acid > *p*-coumaric acid ≥ *m*-coumaric acid, which was confirmed by spectral overlaps of BSA emission spectrum with absorption spectrum of HCA. All free energy changes possessed negative sign indicating the spontaneity of HCA-BSA interaction.**

**Keywords:** Bovine Serum Albumin; Hydroxycinnamic Acid; Fluorescence Quenching; Protein-Ligand Binding

## 1. INTRODUCTION

Recently, considerable attention has been focused on the study of the interaction between small molecules (drugs) and biological macromolecules (e.g. proteins), especially discussing the thermodynamic quality, binding force quality, and mechanism of interactions [1-3]. These studies play crucial role in promoting research on proteins because they can provide useful information for study of pharmacological and biological effects of drugs as well as conformational changes of proteins caused by drugs.

Serum albumin is one of the most abundant proteins in circulatory system of a wide variety of organisms and one of the most extensively studied proteins at all [4,5]. Bovine serum albumin (BSA) consists of 583 amino acids in a single polypeptide chain cross-linked with 17 disulfide bonds. It is composed of three homologous domains (I-III), each of which comprises of two subdomains (A and B). BSA has two tryptophan residues, which significantly contribute to the intrinsic fluorescence of this protein: Trp-134 is located near the surface in domain IB and Trp-212 is buried in a hydrophobic (non-polar) pocket in the internal part of domain IIA [6]. HSA differs from BSA by 24% of primary structure and the most important difference from spectroscopic point of view seems to be presence of only one tryptophan residue (Trp-214) in its molecule [4,5]. Serum albumin possesses a wide range of physiological functions involving the binding, transport and deposition of many endogenous and exogenous ligands present in blood circulation [4,7]. Perhaps, its most outstanding property is the ability to bind a variety of ligands. It is well known that many drugs bind to serum albumin and their effectiveness depends on the binding ability [5,8]. On the other hand, drugs can cause various changes in protein conformation influencing its physiological function and such impaired proteins may be consequently pathologically accumulated in body tissues.

Plant polyphenols represent a heterogeneous group of natural compounds with one or more hydroxyl groups attached to the benzene ring. These substances possess several important physiological roles in plants, such as defense against herbivores and pathogens, pigmentation, and attraction of pollinating insects [9]. The most widely distributed polyphenolic compounds in plant tissues are hydroxycinnamic acids. Some of the most common

naturally occurring HCAs are *p*-coumaric acid, ferulic acid, sinapic acid, and caffeic acid. These can be found in a free form but more often in various conjugated forms resulting from enzymatic hydroxylation, *O*-glycosylation, *O*-methylation or esterification [10,11]. Hydroxycinnamic acids have been reported to possess antimicrobial, antiallergic and anti-inflammatory activities, as well as antimutagenic and immunomodulatory effects [12,13] and due to these properties are widely used in folk medicine. They exert also antioxidant and anti-radical activities [14-16]. Their biological effects are strongly dependent on the number and position of hydroxyl groups [15]. Just the presence of hydroxyl groups suggests the possibility of HCAs binding with molecules of proteins.

Spectroscopic techniques, such as ultraviolet-visible (UV-VIS) absorption spectroscopy [17], fluorescence spectroscopy [1,6], circular dichroism [18], and attenuated total reflectance-Fourier transform infrared spectroscopy [19] are commonly used tools to observe conformational changes in structure of proteins because of non-destructive measurements of substances in low concentration under physiological conditions, high sensitivity, rapidity and ease of implementation. Fluorescence spectroscopy is widely used to study mechanism of the binding between drugs and plasma proteins [6].

Nowadays, the investigation of the binding of naturally occurring polyphenolic compounds with various proteins attracts a great attention. Several spectroscopic studies on the interaction between bovine serum albumin and cinnamic acid [1], ferulic acid [20,21], chlorogenic acid [19,22] or various flavonoids [18,19,24] have been published. Also several studies dealing with the interaction of human serum albumin (HSA) with derivatives of cinnamic acid or flavonoids have been carried out [2,3,17,25-29]. The data obtained in several studies concerning the HCAs-BSA binding parameters (especially caffeic, chlorogenic, and ferulic acid) are hardly comparable because these studies were performed under various conditions (e.g. pH, temperature).

The aim of the presented work was to study interactions of eight naturally occurring hydroxycinnamic acids with bovine serum albumin under physiological conditions (pH 7.4; 37℃) using UV-VIS absorption spectroscopy and fluorescence quenching method, reveal their character, evaluate structure-activity relationships, and compare obtained results with already published spectroscopic data on interaction of HCAs with BSA or HSA. The presented study contributes to the current knowledge in the area of protein-ligand binding, particularly bovine serum albumin-hydroxycinnamic acids interactions.

## 2. EXPERIMENTAL

### 2.1. Chemicals

Bovine serum albumin and all hydroxycinnamic acids were obtained from Sigma-Aldrich GmbH, Germany. The chemical structures of tested HCAs are presented in **Figure 1**. All other chemicals were of analytical grade.

### 2.2. Preparation of Stock Solutions

Bovine serum albumin was dissolved in sodium phosphate buffer (pH 7.4; 0.1 M; 0.05% sodium azide) in order to yield solutions with concentration 16 μM and 2 μM for UV-VIS absorption and fluorescence spectroscopic experiments, respectively. Individual HCAs were dissolved in anhydrous methanol in order to yield 10 mM stock solutions. BSA and HCA solutions were prepared fresh before each measurement.

### 2.3. UV-VIS Absorption Spectroscopy

The UV-VIS spectra were recorded by a spectrophotometer Helios β (Spectronic Unicam, United Kingdom) in a 10 mm quartz cuvette. Quantitative analysis of the potential interaction between HCAs and BSA was performed by the spectroscopic titration. Briefly, solution of BSA (16 μM) was titrated in cuvette by successive additions of HCA solution (10 mM) to a final concentration of 50 μM (the drug to protein molar ratios were 0; 0.25; 0.5; 0.75; 1.0; 1.25; 1.5; 1.75; 2.0; 2.5; and 3.125) and the absorption spectra were recorded from 190 to 550 nm at 37℃.

### 2.4. Fluorescence Spectroscopy

Fluorescence spectra were recorded using a luminescence spectrometer LS-50B (Perkin Elmer, United Kingdom) in a 10 mm quartz Suprasil fluorescence cuvette (Hellma,

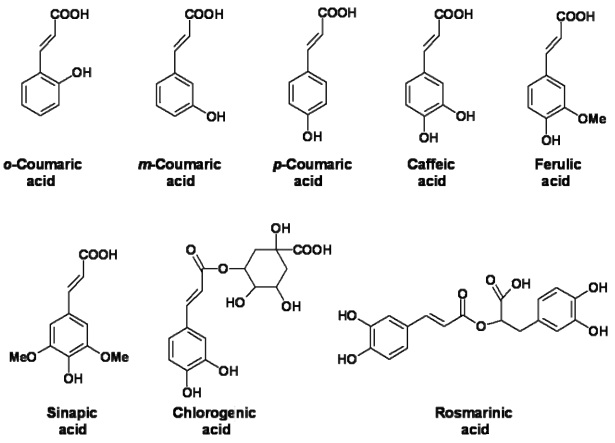



**Figure 1.** Chemical structures of tested hydroxycinnamic acids.

Germany). Fluorescence emission spectra of individual HCA solutions (25 μM) in sodium phosphate buffer (pH 7.4; 0.1 M; 0.05% sodium azide) were recorded. Quantitative analysis of the potential interaction between HCA and BSA was performed by the fluorimetric titration. Briefly, solution of BSA (2 μM) was titrated in cuvette by successive additions of HCA solution (10 mM) to a final concentration of 20 μM (the drug to protein molar ratios were 0; 1.25; 2.5; 3.75; 5.0; 6.25; 7.5; 8.75; and 10.0). Fluorescence emission spectra were recorded from 300 to 530 nm with excitation at 295 nm while stirring. The excitation and emission slits were both set to 5 nm and scanning speed to 200 nm/min. All experiments were carried out at 37℃. Fluorescence intensity was read at emission wavelength of 350 nm.

### 2.5. Principles of Fluorescence Quenching

The intensity of fluorescence can be decreased as a result of a wide variety of processes. Such declines in intensity are called quenching and can be caused by different molecular interactions. Dynamic quenching occurs when the excited-state fluorophore is deactivated upon contact with some other molecule (quencher) and no molecule is chemically altered during this process. In the case of static quenching, a non-fluorescent complex is formed between molecules of fluorophore and quencher. The static and dynamic quenching can be distinguished by the Stern-Volmer analysis. In the case the quenching is either purely static or dynamic, the plot shows linear dependence. When the plot of Stern-Volmer diagram shows exponential dependence, both static and dynamic quenching are exerted [6].

Dynamic quenching of fluorescence is described by the well-known Stern-Volmer equation as follows.

$$F_0 / F = 1 + k_q \tau_0 [Q] = 1 + K_D [Q] \quad (1)$$

In this equation, $F_0$ and $F$ are the fluorescence intensities of BSA in the absence and presence of quencher, respectively, [$Q$] is the quencher concentration, $k_q$ is the bimolecular quenching constant, and $\tau_0$ is the lifetime of the fluorophore in the absence of quencher ($\tau_0$ is about $5.10^{-9}$ s, as to Reference [6]). The Stern-Volmer quenching constant is given by $k_q\tau_0$. In case the quenching is known to be dynamic, the Stern-Volmer constant will be presented by $K_D$ otherwise this constant will be described as $K_S$ [2,6]. The dynamic quenching depends on diffusion, while static quenching does not. One criterion for distinguishing the type of quenching is the fact that the bimolecular quenching constant $k_q$ is larger than diffusion-limited rate constant of the biomolecule ($1 \times 10^{10}$ $M^{-1} \cdot s^{-1}$) [6], so the static mechanism is the main reason that causes the fluorescence quenching (formation of a complex).

When small molecules bind independently to a set of equivalent sites on a macromolecule, the equilibrium between free and bound molecules is given by the equation:

$$\log(F_0 - F)/F = \log K_b + n \log[Q] \quad (2)$$

where $K_b$ represents binding constant for quencher-protein interaction, n the number of binding sites per BSA, and $F_0$, $F$, have the same meaning as in (1) [2]. The values of $K_b$ and n could be determined from the intercept of y-axe and slope by plotting log ($F_0 - F$)/$F$ against log [$Q$]. Utilizing $K_b$, the free energy change ($\Delta G^0$) value can be estimated from the following equation [30]:

$$\Delta G^0 = -RT \ln K_b \quad (3)$$

The negative sing $\Delta G^0$ value confirms the spontaneity of binding.

## 3. RESULTS

### 3.1. Spectroscopic Study of Interactions between BSA and HCAs

The UV-VIS absorption spectra of BSA titrated by individual HCAs solution were monitored in order to explore the structural changes of BSA caused by addition of these compounds. Spectral shifts were observed in all HCA-BSA systems with rising concentration of tested compound. Six HCAs (*p*-coumaric, caffeic, ferulic, sinapic, chlorogenic, and rosmarinic acid) induced move of Trp absorption maximum (280 nm) to longer wavelengths which is called bathochromic (red) shift. Maximal spectral shift was about 6 nm. The opposite phenomenon (blue shift) occurred in the absorption spectrum of BSA after interaction with *o*- or *m*-coumaric acid. Absorbance maximum moved about 5 nm towards shorter wavelengths in both cases (data not shown).

### 3.2. Fluorescence Quenching of BSA in the Presence of HCAs

Quenching of protein intrinsic (tryptophan) fluorescence was employed for more detailed study of HCA-BSA binding. Fluorescence emission spectra were recorded upon excitation at 295 nm, which is attributed to tryptophan residues only. Four individual HCAs (*o*-coumaric, caffeic, sinapic and ferulic acid) possessed remarkable emission maximum at 498, 432, 428 and 414 nm, respectively. The most significant fluorescence intensity showed *o*-coumaric acid. Fluorescence intensities of HCA-BSA systems were read at emission wavelength of 350 nm, where the emission maximum of BSA was located. Protein solution was titrated by successive additions of individual HCA solutions and its fluorescence

intensity gradually decreased with rising concentration of HCA. This may indicate that the microenvironment around tryptophan residues in BSA molecule was altered due to the interaction with tested compound. Fluorescence emission spectrum of *p*-coumaric acid-BSA system is shown in **Figure 2**.

Red shifts of tryptophan emission maximum (350 nm) in dependence on increasing concentration of tested compounds were found in the case of *o*-coumaric, sinapic, chlorogenic, and rosmarinic acid. Emission maximum was slightly shifted towards longer wavelength by 2 nm for both *o*-coumaric and sinapic acid-BSA systems, and by 4 and 5 nm for chlorogenic and rosmarinic acid-BSA system, respectively. Other four tested HCAs did not cause any spectral shift. Emission spectra of *o*-coumaric, caffeic, ferulic, and sinapic acid involved isosbestic point, which might indicate that studied compounds exist both in bound and free form that are in equilibrium. The bound form exerts fluorescence whereas the unbound form does not (**Figure 3**).

It was noticed that emission spectra of these HCA-BSA systems above 430 nm corresponding with emission spectra of individual HCAs.

The type of fluorescence quenching of HCA-BSA systems was distinguished using the Stern-Volmer diagrams in the range of HCA concentrations of 0-20 μM. It was confirmed that the static quenching mechanism is the main reason of protein fluorescence quenching and consecutively the $K_S$ and $k_q$ (1) were determined from the slope of the linear regression curve of $F_0/F$ versus [$Q$] (**Table 1**). The representative Stern-Volmer diagram of *o*-coumaric-BSA system is displayed in the inset of **Figure 3**. Rosmarinic acid exhibited exponential dependence (**Figure 4**) indicating that both types of quenching

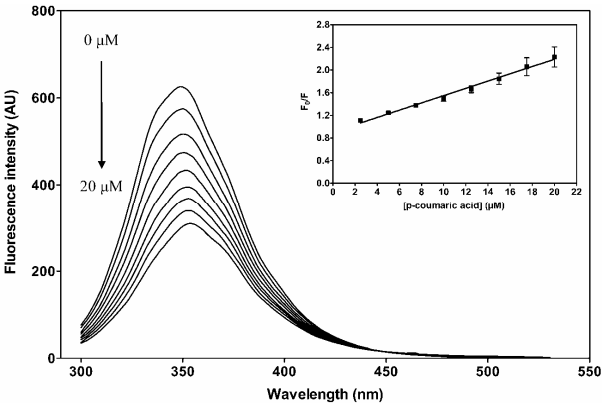



**Figure 2.** Fluorescence emission spectra of BSA (2 μM) in the absence and in the presence of increasing amounts of *p*-coumaric acid (0-20 μM) in sodium phosphate buffer (pH 7.4; 0.1 M; 0.05% sodium azide) at $\lambda_{ex}$ = 295 nm and 37℃. The inset shows the corresponding Stern-Volmer diagram of the *p*-coumaric acid-BSA system ($\lambda_{em}$ = 350 nm), $R^2$ = 0.9908.

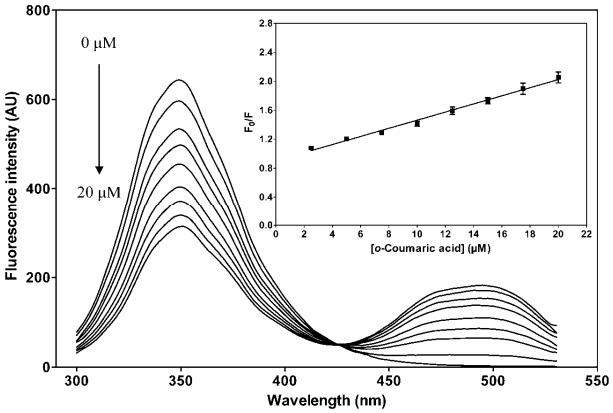



**Figure 3.** Fluorescence emission spectra of BSA (2 μM) in the absence and in the presence of increasing amounts of *o*-coumaric acid (0-20 μM) in sodium phosphate buffer (pH 7.4; 0.1 M; 0.05% sodium azide) at $\lambda_{ex}$ = 295 nm and 37℃. The inset shows the corresponding Stern-Volmer diagram of the *o*-coumaric acid-BSA system ($\lambda_{em}$ = 350 nm), $R^2$ = 0.9921.

**Table 1.** The Stern-Volmer quenching constants ($K_S$) and the bimolecular quenching constants ($k_q$) of the system of HCA-BSA at 37℃.

| Tested compound | $K_S$ ± S.D.[a] [× $10^4$ l.mol$^{-1}$] | $k_q$[b] ± S.D.[a] [× $10^{13}$ M$^{-1}$s$^{-1}$] |
|---|---|---|
| *o*-coumaric acid | 5.95 ± 0.155 | 1.19 ± 0.031 |
| *m*-coumaric acid | 5.96 ± 0.170 | 1.19 ± 0.034 |
| *p*-coumaric acid | 7.13 ± 0.190 | 1.43 ± 0.038 |
| caffeic acid | 4.30 ± 0.263 | 0.86 ± 0.053 |
| ferulic acid | 4.86 ± 0.090 | 0.97 ± 0.018 |
| sinapic acid | 4.25 ± 0.209 | 0.85 ± 0.042 |
| chlorogenic acid | 5.36 ± 0.195 | 1.07 ± 0.039 |

[a]standard deviation (mean value of three independent experiments); [b]$k_q$ = $K_S/\tau_0$; $\tau_0$ = $5.10^{-9}$ s. [6]

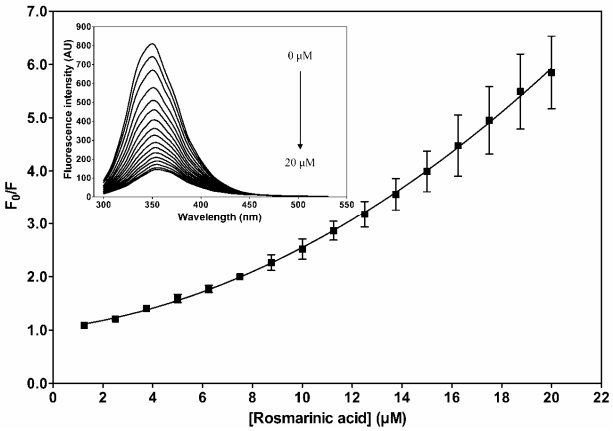



**Figure 4.** The Stern-Volmer diagram of the rosmarinic acid-BSA system obtained by the titration with rosmarinic acid at 37℃. [BSA] = 2 μM, [rosmarinic acid] = 0-20 μM, pH 7.4, $\lambda_{ex}$ = 295 nm, $\lambda_{em}$ = 350 nm. The inset shows the corresponding fluorescence quenching spectra.

were asserted and for this reason the $K_S$ ($k_q$) of the rosmarinic acid-BSA system was not determined.

It was noticed that emission spectra of these HCA-BSA systems above 430 nm corresponding with emission spectra of individual HCAs.

The type of fluorescence quenching of HCA-BSA systems was distinguished using the Stern-Volmer diagrams in the range of HCA concentrations of 0-20 μM. It was confirmed that the static quenching mechanism is the main reason of protein fluorescence quenching and consecutively the $K_S$ and $k_q$ (1) were determined from the slope of the linear regression curve of $F_0/F$ versus [$Q$] (**Table 1**). The representative Stern-Volmer diagram of *o*-coumaric-BSA system is displayed in the inset of **Figure 3**. Rosmarinic acid exhibited exponential dependence (**Figure 4**) indicating that both types of quenching were asserted and for this reason the $K_S$ ($k_q$) of the rosmarinic acid-BSA system was not determined.

### 3.3. Binding Parameters and Binding Mode of BSA-HCA Complexes

Except for the rosmarinic acid-BSA system, the binding constants ($K_b$), binding sites (n), and free energy changes ($\Delta G^0$) of all other HCA-BSA systems have been determined according to the **Eqs. (2)** and **(3)**, respectively. Obtained values are presented in **Table 2** and representative example of binding parameters determination for sinapic acid is displayed in **Figure 5**. The binding affinity was strongest for chlorogenic acid and ranked in the order chlorogenic acid > sinapic acid ≥ caffeic acid > ferulic acid > *o*-coumaric acid > *p*-coumaric acid ≥ *m*-coumaric acid. This order of binding affinities of HCA to BSA was confirmed also by spectral overlaps of BSA emission spectrum with absorption spectrum of individual HCAs. Example of spectral overlap for chlorogenic acid is shown in **Figure 6**. The negative value of $\Delta G^0$ indicating spontaneous process of HCA-BSA binding was determined for all studied interactions (**Table 2**).

**Table 2.** The binding constants ($K_b$), the number of binding sites (n) and the free energy change ($\Delta G^0$) of the HCA-BSA system at 37℃ which showed the static quenching mechanism.

| Tested compound | $K_b$ ± S.D.[a] [× $10^5$ l.mol$^{-1}$] | n ± S.D.[a] | $\Delta G^0$ ± S.D.[a] [kJ.mol$^{-1}$] |
|---|---|---|---|
| *o*-coumaric acid | 3.34 ± 0.720 | 1.17 ± 0.012 | –32.73 ± 0.563 |
| *m*-coumaric acid | 1.31 ± 0.045 | 1.08 ± 0.002 | –30.36 ± 0.088 |
| *p*-coumaric acid | 1.81 ± 0.728 | 1.10 ± 0.027 | –30.98 ± 1.096 |
| caffeic acid | 4.16 ± 1.659 | 1.18 ± 0.011 | –33.12 ± 1.088 |
| ferulic acid | 3.39 ± 0.802 | 1.18 ± 0.019 | –32.75 ± 0.621 |
| sinapic acid | 4.19 ± 0.117 | 1.21 ± 0.004 | –33.36 ± 0.072 |
| chlorogenic acid | 6.67 ± 0.837 | 1.23 ± 0.016 | –34.55 ± 0.325 |

[a]standard deviation (mean value of three independent experiments)

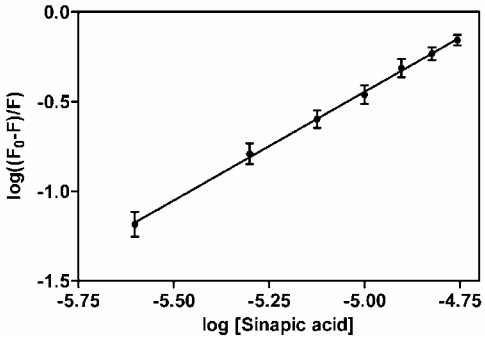



**Figure 5.** Logarithmic plots of fluorescence quenching of BSA treated with different concentrations of sinapic acid at physiological conditions (37℃; pH 7.4). [BSA] = 2 μM, [sinapic acid] = 0-20 μM, $\lambda_{ex}$ = 295 nm and $\lambda_{em}$ = 350 nm. $R^2$ = 0.9988.

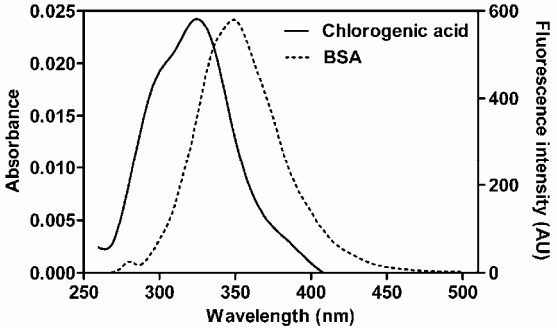



**Figure 6.** Overlap between the fluorescence emission spectrum of BSA and the absorption spectrum of chlorogenic acid at physiological conditions (37℃; pH 7.4). [BSA] = 2 μM, [chlorogenic acid] = 2 μM, $\lambda_{ex}$ = 295 nm and $\lambda_{em}$ = 350 nm.

## 4. DISCUSSION

Red shift in absorption maximum of tryptophan residues indicates changes in its microenvironment, where the polypeptide strand of BSA molecule is less extended and the hydrophobicity around Trp is increased. The conformational stability, rigidity, mechanical strength, and contributions of electrostatic interactions are enhanced by absence of water in the molecular interior [31]. On the other hand, blue shift implies that the BSA polypeptide strands are more extended and the hydrophocibity of Trp vicinity is decreased.

The changes in tryptophan microenvironment polarity are probably related to chemical structure of each HCA, namely to the number and position of hydroxyl groups. Hydrophobicity around tryptophan residues rose with increasing number of hydroxyl groups in the molecule of

HCA. However, the position of hydroxyl groups on the benzene ring seemed to be also important. Only *o*- and *m*-monosubstituted derivatives of cinnamic acid (*o*-coumaric and *m*-coumaric acid) caused increase in the polarity of Trp environment, while *p*-monosubstituted (*p*-coumaric acid), di- and tri-substituted derivatives showed opposite effect.

Emission spectra of BSA were measured using excitation wavelength at 295 nm to ensure that the light caused excitation only of tryptophan residues. These are highly susceptible to any change in their local environment resulting in appearance of a substantial spectral shift [6]. In contrast to absorption spectroscopy, the red shift of Trp emission band is caused by decrease in hydrophobic property of its environment in protein molecule suggesting that tryptophan residue has been brought to more hydrophilic environment [3] and protein secondary structure has been changed [26].

Red shift in emission spectra of four studied hydroxycinnamic acids (*o*-coumaric, sinapic, chlorogenic, and rosmarinic acid), which indicated that binding of these compounds to BSA was associated with changes in the dielectric environment of at least one of its two indole rings, was observed. Only slight blue shift and no shift were noticed during interaction of cinnamic acid with BSA and HSA, respectively [1,2]. It can be expected that changes found in emission spectra of HCA-BSA systems were connected with alterations in microenvironment of Trp-134 because Trp-212 in BSA molecule is in the similar position as Trp-214, where cinnamic acid caused no spectral shift of its emission band. Increasing number of hydroxyl groups in the molecule of HCA was accompanied by decline of hydrophobicity around Trp. The significant changes in BSA emission spectrum were observed after its interaction with chlorogenic or rosmarinic acid. Other authors described red shifts by 3 to 18 nm in HSA spectrum upon reaction with *p*-coumaric, ferulic, sinapic, and chlorogenic acid [2,3,25] but the concentrations applied in these experiments were higher (up to 150 μM) than those used in presented work.

The natural lifetime for the biological macromolecules ($\tau_0$) is generally given as $10^{-8}$ s [1,3]. However, the value for BSA is more precisely estimated as $5 \times 10^{-9}$ s [6]. The latter mentioned value was used in this study for calculation of bimolecular quenching constants ($k_q$) which reflect efficiency of quenching or the accessibility of the fluorophores to the quencher. The fact that the value of $k_q$ is higher than the value of the diffusion-limited rate constant of the biomolecule ($1 \times 10^{10}$ $M^{-1} \cdot s^{-1}$) is one of the criterions for determination of static mechanism of quenching [1,6].

Only small but significant differences among $k_q$ values for the tested HCA-BSA systems were observed. All tested HCAs exerted better quenching effect than cinnamic acid ($2.26 \times 10^{12}$ $M^{-1} \cdot s^{-1}$ for $\tau_0 = 10^{-8}$ s) [1]. Rosmarinic acid caused the most outstanding decrease in fluorescence intensity of BSA in the range of studied concentrations but it exhibited exponential dependence in the Stern-Volmer diagram and thus was not considered in overall comparison. Coumaric acids showed stronger quenching activity than the other more substituted HCAs. This effect was probably dependent upon the position of hydroxyl group on aromatic ring and *p*-position was determined as the most suitable location of hydroxyl group. The $k_q$ value of chlorogenic acid was similar to $k_q$ of coumaric acids and slightly higher than $k_q$ of more substituted derivatives (**Table 1**). It was probably caused by the presence of five hydroxyl groups in its molecule and their spatial arrangement [3]. The lowest quenching effect was observed for sinapic and caffeic acid followed by ferulic acid. Presence of methoxy group seemed to be important for quenching activity too. Ferulic acid showed higher effect than caffeic acid, while the $k_q$ of sinapic acid was slightly lower compared to caffeic acid. This was perhaps caused by steric hindrance in molecule of sinapic acid. The obtained bimolecular quenching constant for chlorogenic acid-BSA system is in good agreement with data found in literature [19]. However, published data are inconsistent and even in one case no quenching by chlorogenic acid was observed [20]. More studies were published for some HCA-HSA systems. The highest $k_q$ was obtained for chlorogenic acid followed by caffeic and sinapic, while ferulic and *p*-coumaric acid possessed lower $k_q$ values. Data obtained for HCA-BSA systems in this study corresponded with these findings by other authors [2,3].

In general, the binding constant $K_b$ reflects the power of ligand-protein association and thus can be used for comparison of binding affinities of structurally-related ligands to protein molecule connected with alterations of its secondary structure. Number of binding sites is another important parameter that contributes to better understanding of ligand-protein interaction [1-3].

The binding constant $K_b$ for cinnamic acid-BSA system mentioned in the literature [1,2] is lower than values obtained in our experiments for HCA-BSA systems, which may confirm significance of hydroxyl groups in the process of binding. Moreover, binding affinity of cinnamic acid is higher for BSA than for HSA, which indicates that also binding of HCAs to BSA may be more pronounced [2]. It was demonstrated that interaction of HCAs with protein molecule depends mainly on the size and structure of ligand, especially on the number and position of hydroxyl groups on the aromatic ring [20,32]. Hydroxyl groups of studied compounds form hydrogen bonds with amino acid residues in the protein

molecule. Another important factor influencing ligand-protein binding is aromaticity of the ligand molecule because hydrophobic interactions are formed between aromatic rings of ligand and amino acid residues [25]. Chlorogenic acid with two aromatic hydroxyls and three hydroxyls on cyclohexane ring exerted the strongest binding affinity because this compound can form hydrogen bonds with protein more easily than other less-substituted HCAs, e.g. ferulic acid. Similar results were published also for ferulic acid-HSA and chlorogenic acid-HSA system [3]. It is evident that the hydroxyls substituted on aromatic ring of the HCAs play an important role in the changes of BSA secondary structure. Monosubstituted HCAs possessed the lowest binding affinities of all HCAs studied. The differences among binding affinities of sinapic, ferulic and caffeic acid were not statistically significant.

The binding parameters of chlorogenic and ferulic acid with bovine serum albumin have been intensively studied by other authors, while no information about other HCAs was found in the literature. The value of $K_b$ reported by Tang *et al.* [19] for chlorogenic acid-BSA system using fluorescence quenching method is lower than value obtained in performed experiments, but the authors used higher concentrations of chlorogenic acid and different experimental conditions. Moreover, Rawel *et al.* [20,21] reported that chlorogenic acid does not quench Trp fluorescence in BSA and determined its binding constant by Hummel-Dreyer/size exclusion chromatography which showed significantly lower value of $K_b$ in comparison with results presented by Tang *et al.* [19]. Non-covalent interactions of chlorogenic acid with BSA have been studied by Prigent *et al.* [33], who reported that these interactions decrease with the increasing temperature while pH and ionic strength had no pronounced effect. Zhang *et al.* [34] reported the $K_b$ for ferulic acid-BSA system determined by affinity capillary electrophoresis which is in good agreement with our result. By contrast, Rawel *et al.* [20] determined binding constant of ferulic acid-BSA system by fluorescence quenching method and Hummel-Dreyer/size exclusion chromatography, where first method gave similar results to our data and $K_b$ obtained by second method was significantly lower. However, several studies dealing with binding of *p*-coumaric, caffeic, ferulic, sinapic acid, and chlorogenic acid with HSA have been published [2,3,25].

The results showed that the numbers of binding sites ranged between 1.08 and 1.23 suggesting that one molecule of BSA was associated with one molecule of HCA in the drug to protein ratio up to 10 for the tested HCAs apart. The number of binding sites rose with increasing number of hydroxyl groups in the ligand molecule.

The free enthalpy had negative sign for all studied interactions which indicates the spontaneity of the interaction between BSA and hydroxycinnamic acids. The relatively strong binding enthalpy underlines the stability of BSA-HCA complexes from the energetic point of view. These findings are supported by data found in literature [1,20,21].

## 5. CONCLUSIONS

Apart from rosmarinic acid, all tested HCAs quenched tryptophan fluorescence of BSA in the studied range of concentrations (0-20 μM) mainly by static quenching mechanism and thus showed the formation of non-fluorescent HCA-BSA complexes. For this reason the rosmarinic acid-BSA system was not concluded in the overall assessment of binding affinities. The obtained results suggest that the binding affinity and number of binding sites depend on the number and position of hydroxyl groups in the molecule of HCA. Disubstituted and trisubstituted derivatives exhibited stronger binding affinity than monosubstituted derivatives. The number of binding sites for all HCAs ranged from 1.08 to 1.23 suggesting that one molecule of BSA associates with one molecule of HCA. All HCA-BSA interactions were spontaneous processes based on $\Delta G^0$. The results imply that HCAs could be stored and transported in blood by serum albumin which may influence their biological and pharmacological activities. On the other hand, physiological functions of this protein could be altered by ligand binding.

## 6. ACKNOWLEDGEMENTS

The presented study was supported by the Specific research 2009 of the University of Hradec Králové.

## REFERENCES

[1] Bian, H., Zhang, H., Yu, Q., Chen, Z. and Liang, H. (2007) Studies on the interaction of cinnamic acid with bovine serum albumin. *Chemical & Pharmaceutical Bulletin*, **55(6)**, 871-875.

[2] Jiang, M., Xie, M.X., Zheng, D., Liu, Z., Li, X.Y. and Chen, X. (2004) Spectroscopic studies on the interaction of cinnamic acid and its hydroxyl derivatives with human serum albumin. *Journal of Molecular Structure*, **692(1-2)**, 71-80.

[3] Kang, J., Liu, Y., Xie, M.X., Li, S., Jiang, M. and Wang, Y.D. (2004) Interactions of human serum albumin with chlorogenic acid and ferulic acid. *Biochimica et Biophysica Acta*, **1674(2)**, 205-214.

[4] Peters, T. (1996) All about albumin: Biochemistry, genetics, and medical applications. Academic Press, San Diego.

[5] Carter, D.C. and Ho, J.X. (1994) Structure of serum albumin. *Advances in Protein Chemistry*, **45**, 153-203.
[6] Lakowicz, J. R. (2004) Principles of fluorescence spectroscopy. 2nd Edition, Springer, New York.
[7] Behrens, P.Q., Spiekerman, A.M. and Brown, J.R. (1975) Structure of bovine serum-albumin. *Federation Proceedings*, **34(3)**, 591.
[8] Kragh-Hansen, U. (1981) Molecular aspects of ligand binding to serum albumin. *Pharmacological Reviews*, **33(1)**, 17-53.
[9] Dewick, P. M. (2002) Medicinal natural products: A biosynthetic approach. 2nd Edition, Wiley, Chichester.
[10] Harborne, J.B. and Williams, C.A. (2000) Advances in flavonoid research since 1992. *Phytochemistry*, **55(6)**, 481-504.
[11] Rice-Evans, C.A., Miller, N. and Paganga, G. (1997) Antioxidant properties of phenolic compounds. *Trends in Plant Science*, **2(4)**, 152-159.
[12] Natarajan, K., Singh, S., Burke, T.R., Grunberger, D. and Aggarwal, B.B. (1996) Caffeic acid phenethyl ester is a potent and specific inhibitor of activation of nuclear transcription factor NF-kappa B. *Proceedings of the National Academy of Sciences*, USA, **93(17)**, 9090-9095.
[13] Pannala, A.S., Razaq, R., Halliwell, B., Singh, S. and Rice-Evans, C.A. (1998) Inhibition of peroxynitrite dependent tyrosine nitration by hydroxycinnamates: Nitration or electron donation? *Free Radical Biology and Medicine*, **24(4)**, 594-606.
[14] Natella, F., Nardini, M., Di Felice, M. and Scaccini, C.J. (1999) Benzoic and cinnamic acid derivatives as antioxidants: Structure-activity relation. *Journal of Agricultural and Food Chemistry*, **47(4)**, 1453-1459.
[15] Rice-Evans, C.A., Miller, N.J. and Paganga, G. (1996) Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free Radical Biology and Medicine*, **20(7)**, 933-956.
[16] Graf, E. (1992) Antioxidant potential of ferulic acid. *Free Radical Biology and Medicine*, **13(4)**, 435-448.
[17] Kanakis, C.D., Tarantilis, P.A., Polissiou, M.G., Diamantoglou, S. and Tajmir-Riahi, H.A. (2006) Antioxidant flavonoids bind human serum albumin. *Journal of Molecular Structure*, **798(1-3)**, 67-74.
[18] Tian, J., Liu, J., Hu, Z. and Chen, X. (2005) Interaction of wogonin with bovine serum albumin. *Bioorganic & Medicinal Chemistry*, **13(12)**, 4124-4129.
[19] Tang, D., Li, H.J., Wen, X.D. and Qian, Z.M. (2008) Interaction of bioactive components caffeoylquinic acid derivatives in chinese medicines with bovine serum albumin. *Chemical & Pharmaceutical Bulletin*, **56(3)**, 360-365.
[20] Rawel, H.M, Frey, S.K., Meidtner, K., Kroll, J. and Schweigert, F.J. (2006) Determining the binding affinities of phenolic compounds to proteins by quenching of the intrinsic tryptophan fluorescence. *Molecular Nutrition & Food Research*, **50(8)**, 705-713.
[21] Rawel, H.M, Meidtner, K. and Kroll, J. (2005) Binding of selected phenolic compounds to proteins. *Journal of Agricultural and Food Chemistry*, **53(10)**, 4228-4235.
[22] Rawel, H.M, Rohn, S., Kruse, H.P. and Kroll, J. (2002) Structural changes induced in bovine serum albumin by covalent attachment of chlorogenic acid. *Food Chemistry*, **78(4)**, 443-455.
[23] Tian, J., Liu, J., Tian, X., Hu, Z. and Chen, X. (2004) Study of the interaction of kaempferol with bovine serum albumin. *Journal of Molecular Structure*, **691(1-3)**, 197-202.
[24] Papadopoulou, A., Green, R.J. and Frazier, R.A. (2005) Interaction of flavonoids with bovine serum Albumin: A fluorescence quenching study. *Journal of Agricultural and Food Chemistry*, **53(1)**, 158-163.
[25] Liu, Y., Xie, M.X., Jiang, M. and Wang, Y.D. (2005) Spectroscopic investigation of the interaction between human serum albumin and three organic acids. *Spectrochimica Acta Part A*, **61(9)**, 2245-2251.
[26] He, W., Li, Y., Xue, C., Hu, Z., Chen, X. and Sheng, F. (2005) Effect of Chinese medicine alpinetin on the structure of human serum albumin. *Bioorganic & Medicinal Chemistry*, **13(5)**, 1837-1845.
[27] Xie, M.X., Long, M., Liu, Y., Qin, C. and Wang, Y.D. (2006) Characterization of the interaction between human serum albumin and morin. *Biochimica et Biophysica Acta*, **1760(8)**, 1184-1191.
[28] Xie, M.X., Xu, X.Y. and Wang, Y.D. (2005) Interaction between hesperetin and human serum albumin revealed by spectroscopic methods. *Biochimica et Biophysica Acta*, **1724(1-2)**, 215-224.
[29] Liu, J., Tian, J., Li, Y., Yao, X., Hu, Z. and Chen, X. (2004) Binding of the bioactive component daphnetin to human serum albumin demonstrated using tryptophan fluorescence quenching. *Macromolecular Bioscience*, **4(5)**, 520-525.
[30] Shang, L., Jiang, X. and Dong, S. (2006) In vitro study on the binding of neutral red to bovine serum albumin by molecular spectroscopy. *Journal of Photochemistry and Photobiology A*, **184(1-2)**, 93-97.
[31] Hu, Y.J., Liu, Y., Zhao, R.M., Dong, J.X. and Qu, S.S. (2006) Spectroscopic studies on the interaction between methylene blue and bovine serum albumin. *Journal of Photochemistry and Photobiology A*, **179(3)**, 324-329.
[32] Bartolome, B., Estrella, I. and Hernandez, M.T. (2000) Interaction of low molecular weight phenolics with proteins (BSA). *Journal of Food Science*, **65(4)**, 617-621.
[33] Prigent, S.V.E., Gruppen, H., Visser, A.J.W.G., van Koningsveld, G.A., de Jong, G.A.H. and Voragen, A.G.J. (2003) Effects of non-covalent interactions with 5-O-caffeoylquinic acid (chlorogenic acid) on the heat denaturation and solubility of globular proteins. *Journal of Agricultural and Food Chemistry*, **51(17)**, 5088-5095.
[34] Zhang, Y., Xu, M., Du, M. and Zhou, F. (2007) Comparative studies of the interaction between ferulic acid and bovine serum albumin by ACE and surface plasmon resonance. *Electrophoresis*, **28(11)**, 1839-1845.

**Vol.2, No.6, 571-575 (2010)**
doi:10.4236/ns.2010.26072

**Natural Science**

# Evolution of *Homo sapiens* in Asia: an alternative implication of the "Out-of-Africa" model based on mitochondrial DNA data

**Hiroto Naora**

Research School of Biology, The Australian National University, Canberra, Australia; hiroto.naora@anu.edu.au

Received 25 September 2009; revised 9 November 2009; accepted 14 January 2010.

## ABSTRACT

**Cann *et al.* [1] have claimed on the basis of mitochondrial DNA (mtDNA) data that our direct ancestral *Homo sapiens* evolved in the African continent and spread to other continents, followed by the total replacement of the indigenous population. Their "Out-of-Africa" model is based on the assumption that mtDNA inheritance is simply maternal. Recent findings suggest the possibility that in between-population, e.g. African and Asian, mating, the African paternal mtDNA was transferred to the egg cell of an Asian together with Y-chromosomal DNA in the human past. Considering that Y-chromosomal DNA and mtDNA sequences of African origin coexist together with Asian X-chromosomal and autosomal DNA sequences in a current Asian, the observations by Cann *et al.* suggest the full/near full replacement of mtDNA in the human past, but do not necessarily imply the total replacement of indigenous populations with African migrants.**

**Keywords:** Out-of-Africa Model; Origin of Asian; Paternal mtDNA Inheritance; mtDNA Transmission/ Recombination

## 1. INTRODUCTION

Evidence has accumulated that the *Homo* lineage originally appeared in Africa, followed by its successful global expansion. The view of "Out-of-Africa" that our direct ancestral *H. sapiens* evolved in the African continent and spread to other continents, has been popularly received among researchers [1-8]. On the other hand, there have been significant fossil records in non-African continents, supporting the "regional continuity" model [9]. This model claims that our direct ancestral *H. sapiens* evolved locally in the widespread regions of major continents, e.g. Africa, Europe and Asia [10-12]. In fact, morphological continuity in East Asian traits from East Asian *Homo erectus* during the middle-late Pleistocene transition can be seen in these fossil records [12-14]. However, most of these records have neither definitely refuted nor supported one of the models for or against the origin of *H. sapiens*, particularly in Asia. Furthermore, most genetic evidence, such as simulated dendrograms, genetic diversity and ancient DNA sequences can argue for either model of human origin [15].

In 1987, a crucial observation was made by Cann *et al.*, who examined the human mitochondrial DNA (mt DNA) diversity of globally dispersed populations. They concluded that modern humans simply spread to other continents, e.g. East Asia, from Eastern Africa around 200,000 years to 50,000 years (200 KY to 50 KY) ago, followed by the total replacement of pre-inhabited indigenous populations [1,2]. The "Out-of-Africa" model, reconstructed with mtDNA data, is primarily based on a few assumptions. One of the key assumptions is that animal mtDNA does not undergo homologous recombination. This is because of past failure to observe the clear cases of mtDNA recombination in natural populations. However, recent findings have raised a new insight into mtDNA inheritance and the behaviours of mtDNA in a fertilised egg [16,17]. In this study, an attempt was made to integrate the interdisciplinary information obtained in the research fields, not only anthropology and mitochondrial genetics but also other areas, such as developmental biology, ecology and social sciences. A meta-analysis of these findings has raised the need of careful scrutiny in the interpretation of Cann *et al.* [1,2].

## 2. CRITICAL VIEW OF MITOCHONDRIAL DNA INHERITANCE: STRICTLY MATERNAL?

Two sibling bat species, European *Myotis myotis* and newly migrated *M. blythii* to Europe from Asia, share

multiple identical or very similar haplotype in mitochondrial genomes when they occur in sympasy. However, they show a strikingly different pattern of nuclear DNA diversity. On the other hand, allopatric *M. blythii* in Asia possesses mtDNA divergent from those of two species in Europe, postulating that the mtDNA of European *M. blythii* has been replaced with that of *M. myotis* [18]. Similar mtDNA (full/near-full/partial) replacement has been observed in a wide variety of other species in an animal kingdom, for example, nematodes [19], molluscs [20], insects [21], fishes [22] and mammals other than bats, such as vole [23]. As opposed to the popular belief, all of these observations suggest that paternal mtDNA inheritance occurs widely in the animal kingdom. In fact, paternal mtDNA inheritance has been observed even in humans [24,25]. At present, no specific information has been reported, which strongly favours the notion that these human cases were really exceptional and that human mtDNA replacement, similar to the cases of *Myotis*, has never taken place in evolution when two "sibling" human populations met and lived in sympasy. Furthermore, evidence has shown that the segregation and transmission of mtDNA sequences [26-28] play a crucial role in the inheritance of human mitochondrial diseases [16].

Molecular and cellular events at an early stage of fertilisation would also provide us with more information which we can not ignore the view of paternal mtDNA inheritance in human. During the process of fertilisation in mammals, up to 100 sperm mitochondria actually enter an egg cell, together with the sperm head [17,29, 30]. However, these paternal mtDNAs are soon destroyed by a selective elimination mechanism, including ubiquitination of paternal mitochondria [30]. However, the elimination mechanism in the fertilised egg cell is not always highly stringent and fails to destroy all of the inserted paternal mitochondria, causing paternal leakage [30]. Such leakage was observed in mice progeny at a frequency of around one in 10,000 [31]. An interesting observation was that when "non-self" paternal mitochondria—in terms of the population/subspecies/species, to which the mating partner belongs—are inserted into the egg cell, such as in the mating between cow and gaur, these mitochondria are not eliminated and remain active in the fertilised egg cell [30]. In the mating between two inbred mouse strains, *Mus musculus* and *M. spretus*, a similar escape of paternal mitochondria through the elimination mechanism was observed as well [32]. These leakages would lead to a heteroplasmic offspring and thereby enhance the opportunity of transmission or recombination. It has been shown that human mitochondrial particles are fully equipped with the toolkits required for recombination [33-35]. Furthermore, each mitochondrial organelle holds the mechanisms, leading from heteroplasmy to the transmission of mtDNA [16, 28]. The detailed molecular mechanisms for paternal mtDNA inheritance and replacement might differ from one case to the other [36] and, at present, remain to be examined in each case.

## 3. THE POSSIBLE FULL/NEAR-FULL REPLACEMENT OF MTDNA SEQUENCES IN THE HUMAN PAST

It is highly likely that the anatomically modern humans that originated in Africa were different from those who inhabited in allopasy in the Asian continent for a long evolutionary period [37]. Furthermore, different populations have differing variation in biological responses [38]. Therefore, African migrants and indigenous inhabitants in Asia were likely to show different recognition responses to the partner's paternal mitochondria in their "between-population" mating although they were sibling and hybridisable. Thus, the ubiquitin-tagged paternal mitochondria, which were inserted in the egg cell, would not be subject to stringent mitochondrial elimination and could survive in the fertilised egg cell, as seen in the case of cow and gaur pairing [30]. There would be a huge difference in recognition response between Palaeolithic African migrant and inhabited Asians, who had never exposed to Africans and thus much more paternal (African) mtDNA molecules might had remained active without any damages in Asian egg cells in their "between-population" mating. Since current human populations have already mixed each other in some degree, I believe that much more African paternal mtDNA molecules had survived in Palaeolithic Asian egg cells than those we suspect in current "between-population" mating.

Most, if not all, of advocates for the "Out-of-Africa" model based on mtDNA data often argue against the alternative model of human origine on the following basis: The alternative model has not based on the conclusive evidence showing the recombination/paternal transmission of mtDNA in the human past. However, we should realise that the "Out-of-Africa" model has been standing on the shaky—recently much more shaky—ground without any explicit and conclusive evidence, showing that any recombination/paternal transmission of mtDNA had not taken place in the human past. In fact, Wilson [39] has mentioned that many puzzles have remained between Y-chromosome and mtDNA data in the conservative interpretation of the complex data in human migration. In the present paper, I have already mentioned the new findings, which include the paternal mtDNA behaviours in the fertilised eggs and the recombination/paternal transmission of mtDNA in the wide

range of animal kingdom. In fact, taking all of these new findings into consideration, it is much more difficult to prepare a reasonable explanation for the notion that a series of the events, leading to the replacement of maternal mtDNA with the paternal DNA, had never taken place in the human past. As already mentioned, the transmission or recombination of mtDNA in the animal kingdom has occurred more frequently and more widely than we have previously suspected. Therefore, the limited human cases that have been currently reported so far on the replacement of mtDNA should not be claimed against the argument for the possible replacement events in the human past.

Cann *et al.* [1] have shown that the mtDNA itself of current humans in the Asian continent is really of African origin. However, their claim that indigenous inhabitants in Asia became completely extinct and were totally replaced with African migrants is confusing. As will be discussed in the next section, the present meta-analysis shows that ironically, their result together with chromosomal DNA sequences can nicely account for the possible paternal mtDNA inheritance in the human past. Thus, this novel view does not necessarily imply the extinction of Asian indigenous inhabitants, followed by the *total* replacement of the human population in Asia.

## 4. DISCUSSION AND CONCLUSION

The estimated ages of the most recent common ancestor (MRCA) of Y-chromosomal DNA sequences, such as several sites on SRY and YAP regions, were around 150 KY ago [40]. Current non-African men carry the M168 mutation, which arose in Africa during the period of 89 KY to 35 KY ago [41-43]. All of these sequences on Y-chromosome were much younger than those (1,860 KY to 535 KY ago) of X-chromosomal DNA sequences, e.g. gene coded for pyruvate dehydrogenase E1α [44] and non-coding sequences Xq [45], and autosomal DNA sequences, e.g. gene coded for β-globin [46] and non-coding sequences on chromosome 22 [47]. It should be noted that the MRCA ages of Y-chromosomal DNA sequences roughly correspond to the time (200 KY-100 KY ago) of human migration to Asia and thus that the Y-chromosomal and mtDNA sequences were likely to arise in the ancestor who lived in Africa around < 200 KY ago [2]. On the other hand, the DNA sequences on X-chromosome and autosomes could be traced back to the era of *H. erectus* in Asia [10,48-50]. Therefore, these results clearly suggest that the current Asians would be the offspring of the hybrids resulting from the mating between migrated Africans and indigenous Asian inhabitants. However, it appears likely that African migrants brought only African Y-chromosomal and mt DNA sequences to Asia [51] and most of Asian X-chromosomal and autosomal DNA sequences remained in the hybrid offspring. The most plausible scenario of the event in Palaeolithic Asia would be as follows: In the "between-population" mating, African Y-chromosomal DNA entered an Asian egg cell, accompanying with his mtDNA. The newly inserted paternal mtDNA remained intact without any significant damage/elimination in the fertilised egg cell and then formed heteroplasmy in the hybrid. After a series of transmission or recombination processes, the maternal mtDNA of Asian origin was fully/near-fully replaced with the paternal (African) mtDNA. Considering the possibility that the dilution-out and/or selective sweep of African X-chromosomal and autosomal DNA sequences might have occurred for evolutionary advantage, the X-chromosomal and autosomal DNA sequences of Asian origin would tend to be preserved more often and finally would be maintained as a major component in the East Asian population [52]. Therefore, the male offspring, including a current Asian male, would not necessarily be the direct descendant of a replaced African, but would be the hybrid offspring possessing the mitochondrial and Y-chromosomal DNA sequences of African origin, together with Asian X-chromosomal and autosomal DNA sequences. The sequence data, if possible, of chromosomal DNA and mtDNA of Palaeolithic Asian remains and the comparison with those of current humans in different populations should give a brighter view on this issue. The view raised in this paper would open a new research field of biological interactions, particularly in "between-population" reproduction in the human past.

## 5. ACKNOWLEDGEMENTS

I thank Prof. D. Clark-Walker (The Australian National University) and Dr. A. Thorne (The Australian National University) for their critical reading of the manuscript, and valuable suggestions and conversations.

## REFERENCES

[1] Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) Mitochondrial DNA and human evolution. *Nature*, **325 (6099)**, 31-36.

[2] Cann, R. (2001) Genetic clues to dispersal in human populations: Retracing the past from the present. *Science*, **291(5509)**, 1742-1748.

[3] Ingman, M., Kaessmann, H., Pääbo, S., *et al.* (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408(6813)**, 708-713.

[4] Stringer, C. (2003) Human evolution: Out of Ethiopia. *Nature*, **423(6941)**, 692-695.

[5] White, T., Asfaw, B., de Gusta, D., *et al.* (2003) Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature*, **423(6941)**, 742-747.

[6] McDougall, I., Brown, F. and Fleagll, J. (2005) Stratigraphic placement and age of modern humans from Kish, Ethiopia. *Nature*, **433(7027)**, 733-736.
[7] Manica, A., Amos, W., Balloux, F. *et al.* (2007) The effect of ancient population bottle necks on human phenotypic variation. *Nature*, **448(7151)**, 346-349.
[8] Gibbons, A. (2009) Africans' deep genetic roots reveal their evolutionary story. *Science*, **324(5927)**, 575.
[9] Thorne, A.G. and Wolpoff, M.H. (1991) Conflict over modern human origins. *Search*, **22**, 175-177.
[10] Brooks, A.S. and Wood, B. (1990) The Chinese side of the story. *Nature*, **344(6264)**, 288-289.
[11] Frayer, D.W., Wolpoff, M.H., Thorne, A.G., *et al.* (1993) Theories of modern human origins: The palaeontological test. *American Anthropologist*, **95(1)**, 73-96.
[12] Shang, H., Tong, H., Zhang, S., *et al.* (2007) An early modern human from Tianyuan Cave, Zhoukoudian, China. *Proceedings of the National Academy of Sciences*, USA, **104(16)**, 6573-6578.
[13] Etler, D.A. (1994) The Chinese Hominidae: New findings, new interpretations. The Ph. D. thesis, submitted to the Graduate Division of the University of California at Berkeley.
[14] Wu, X. and Poirier, F.E. (1995) Human evolution in China: A metric description of the fossils and a review of the sites. Oxford University Press, New York.
[15] Eller, E., Hawks, J. and Relethford, J.H. (2004) Local extinction and recolonization, species effective population size and modern human origins. *Human Biology*, **76(5)**, 689-709.
[16] Howell, N., Chinnery, P.F., Ghosh, S.S., *et al.* (2000) Transmission of the human mitochondrial genome. *Human Reproduction*, **15(Suppl 2)**, 235-245.
[17] Rokas, A., Ladoukakis, E. and Zouros, E. (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution*, **18(8)**, 411-417.
[18] Berthier, P., Excoffier, L. and Ruedi, M. (2006) Reccurrent replacement of mtDNA and cryptic hybridization between two sibling species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B*, **273(1605)**, 3101-3109.
[19] Lunt, D.H. and Hyman, B.C. (1997) Animal mitochondrial DNA recombination. *Nature*, **387(6630)**, 247.
[20] Ladoukakis, E.D. and Zouros, E. (2001) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution*, **18(7)**, 1168-1175.
[21] Powell, J.R. (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: Evidence from *Drosophila*. *Proceedings of the National Academy of Sciences*, USA, **80(2)**, 492-495.
[22] Bernatchez, L., Glemet, H., Wilson, C.C., *et al.* (1995) Introgression and fixation of arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fishery and Aquatic Sciences*, **52(1)**, 179-185.
[23] Telgelström, H. (1987) Transfer of mitochondrial DNA from northern red-backed vole (*Clethrionomys rutilus*) to bank vole (*C. glareolus*). *Journal of Molecular Evolution*, **24(3)**, 218-227.
[24] Hagelberg, E., Goldman, N., Lió, P., *et al.* (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proceedings of the Royal Society B*, London, **266(1418)**, 485-492.
[25] Schwartz, M. and Visssing, J. (2002) Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine*, **347(8)**, 576-580.
[26] Birky, Jr, C.W., Acton, A.R., Dietrich, R., *et al.* (1982) Mitochondrial transmission genetics: replication, recombination and segregation of mitochondrial DNA and its inheritance in crosses. In: Stonimski, P., Borst, P. and Attardi, G., Eds., *Mitochondrial genes*. Cold Spring Harbor Laboratory, Cold Spring Harbor, 333-348.
[27] Hauswirth, W.W. and Laipis, P.J. (1982) Rapid variation in mammalian mitochondrial genotypes: Implications for the mechanism of maternal inheritance. In: Stonimski, P., Borst, P. and Attardi, G., Eds., *Mitochondrial genes*. Cold Spring Harbor Laboratory, Cold Spring Harbor, 137-141.
[28] Hauswirth, W.W. and Laipis, P.J. (1985) Transmission genetics of mammalian mitochondria: A molecular model and experimental evidence. In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M., Eds., *Achievement and perspectives of mitochondrial research. Vol. 2: Biogenesis*, Elsevier Scientific Publishers, Amsterdam, 49-59.
[29] Ankel-Simons, F. and Cummins, J.M. (1996) Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proceedings of the National Academy of Sciences*, USA, **93(4)**, 13859-13863.
[30] Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., *et al.* (2000) Ubiquitinated sperm mitochondria, selective proteolysis and the regulation of mitochondrial inheritance in mammalian embryo. *Biology of Reproduction*, **63(2)**, 582-590.
[31] Gyllensten, U., Wharton, D., Josefsson, A. *et al.* (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature*, **352(6332)**, 255-257.
[32] Shibata, H., Hayashi, J.-I., Takahama, S., *et al.* (1998) Maternal inheritance of mouse mtDNA in interspecific hybrids: Segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics*, **148(2)**, 851-857.
[33] Thyagarajan, B., Padua, R.A. and Campbell, C. (1996) Mammalian mitochondria possess homologous DNA recombination activity. *Journal of Biological Chemistry*, **271(44)**, 27536-27543.
[34] Lakshmipathy, U. and Campbell, C. (1999) Double strand break rejoining by mammalian mitochondrial extracts. *Nucleic Acids Research*, **27(4)**, 1198-1204.
[35] Lakshmipathy, U. and Campbell, C. (1999) The human DNA ligases III gene encodes nuclear and mitochondrial proteins. *Molecular and Cellular Biology*, **19(5)**, 3869-3876.
[36] Zhao, X., Li, N., Guo, W., *et al.* (2004) Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity*, **93(4)**, 399-403.
[37] Oppenheimer, S. (2003) Out of Eden. The peopling of the world. Constable and Robinson, London.
[38] Ashcroft, R. (2006) Race in medicine: From probability to categorical practice. In: Ellison, G.T.H. and Goodman, A.H., Eds., The nature of difference: Science, society and human biology. CRC press, Boca Raton, 135-153.
[39] Wilson, J.A.P. (2008) A new perspective on later migra-

tion(s). The possible recent origin of some native American halotypes. *Critique of Anthropology*, **28(3)**, 267-278.

[40] Hammer, M.F., Karafet, T., Rasanayagam, A., *et al.* (1998) Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. *Molecular Biology and Evolution*, **15(4)**, 427-441.

[41] Underhill, P.A., Shen, P., Lin, A.A., *et al.* (2000) Y chromosome sequence variation and history of human population. *Nature Genetics*, **26(3)**, 358-361.

[42] Ke, Y., Su, B., Song, X., *et al.* (2001) African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes. *Science*, **292(5519)**, 1151-1153.

[43] Balter, M. (2001) Anthropologists duel over modern human origins. *Science*, **291(5509)**, 1728-1729.

[44] Harris, E. and Hey, J. (1999) X-chromosome evidence for ancient human histories. *Proceedings of the National Academy of Sciences*, USA, **96(6)**, 3320-3324.

[45] Kaessmann, H., Heissig, F., von Haeseler, A., *et al.* (1999) DNA sequence variation in non-coding region of low recombination on the human X-chromosome. *Nature Genetics*, **22(1)**, 78-81.

[46] Harding, R.M., Fullerton, S.M., Griffiths, R.C., *et al.* (1997) Archaic African and Asian lineage in the genetic ancestry of modern humans. *American Journal of Human Genetics*, **60(4)**, 772-789.

[47] Zhao, Z., Jin, L., Fu, Y.-X., *et al.* (2000) World wide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proceedings of the National Academy of Sciences*, USA, **97(21)**, 11354-11358.

[48] Zhu, R.X., Potts, R., Xie, F., *et al.* (2004) New evidence on the earliest human presence at high northern latitudes in northeast Asia. *Nature*, **430(6999)**, 559-566.

[49] Dennell, R. and Roebroeks, W. (2005) An Asian perspective on early human dispersal from Africa. *Nature*, **438(7071)**, 1099-1104.

[50] Shen, G., Gao, X., Gao, B., *et al.* (2009) Age of Zoukoudian *Homo erectus* determined with $^{26}Al/^{10}B$ burial dating. *Nature*, **458(7235)**, 198-200.

[51] Zhang, F., Su, B., Zhang, Y.-P., *et al.* (2007) Genetic studies of human diversity in East Asia. *Philosophical Transactions of Royal Society B*, **362(1482)**, 987-995.

[52] Naora, H. (2007) Morphological variation and sexual behaviour in the human past. II. The origin of East Asians and their sexual behaviour. *Dokkyo Journal of Medical Sciences*, **34(2)**, 141-151.

**Vol.2, No.6, 576-589 (2010)**
doi:10.4236/ns.2010.26073

**Natural Science**

# Building reliable genetic maps: different mapping strategies may result in different maps

**Yefim Ronin, David Mester, Dina Minkov, Abraham Korol***

Institute of Evolution and Department of Evolutionary and Environmental Biology University of Haifa, Mount Carmel, Haifa, Israel;
*Corresponding Author: korol@research.haifa.ac.il

Received 24 November 2009; revised 4 January 2010; accepted 6 April 2010.

## ABSTRACT

**New high throughput DNA technologies resulted in a disproportion between the high number of scored markers for the mapping populations and relatively small sizes of the genotyped populations. Correspondingly, the number of markers may, by orders of magnitude, exceed the threshold of recombination resolution achievable for a given population size. Hence, only a small part of markers can be genuinely ordered in the map. The question is how to choose the most informative markers for building such a reliable "skeleton" map. We believe that our approach provides a solution to this difficult problem due to: 1) powerful tools of discrete optimization for multilocus ordering; 2) a verification procedure, which is impossible without fast and high-quality optimization, to control the map quality based on re-sampling techniques; 3) an interactive algorithm of marker clustering in complicated situations caused by significant deviation of recombination rates between markers of non-homologous chromosomes from the expected 50% (referred to as quasi-linkage or pseudo-linkage); and 4) an algorithm for detection and removing excessive markers to increase the stability of multilocus ordering.**

**Keywords:** Pseudo-Linkage; Skeleton Map; Map Verification; Map Stability; Traveling Salesperson Problem; Guided Evolution Strategy

## 1. INTRODUCTION

Genetic maps are an important tool in genomics and in numerous practical applications such as breeding, medical genetics, and gene cloning. Unfortunately, the available algorithms and software tools become less suitable with an increasing number of available markers. The objective of our study is to develop efficient methodology for building multilocus genetic maps, providing the control of the quality of maps by detecting and removing the sources of map instability. Two major problems should be addressed in multilocus genetic mapping: 1) markers that belong to non-homologous chromosomes should not be assigned to the same linkage group; and 2) markers from the same chromosome should be placed on the genetic map in the same order as the corresponding DNA sequences that reside in the chromosome.

In situations with significant deviations of the recombination rates between non-synthetic markers from the expected level (50%), the problem of correct clustering cannot be solved by an arbitrary choice of a certain (constant) threshold value of recombination or LOD, albeit this is exactly how this problem is treated in many mapping packages [1,2]. Indeed, in experiments with the foregoing characteristics, the recombination values between groups of markers from different chromosomes may sometimes be smaller than the values between adjacent markers within a chromosome. This phenomenon, referred to as "quasi-linkage" (or "pseudo-linkage") can result from a combination of statistical and biological reasons and scoring errors. The statistical reasonsof pseudo-linkage are mainly caused by the sample size and number of chromosomes ($n$) in the genome: increased $n$ is associated with higher chances to detect "significant" deviations from independent segregation. But the major source of pseudo-linkage is biology. Literature on this phenomenon in many species can be found in: [3-5] and references therein.

Mapping algorithms tend to ignore pseudo-linkage. Consequently, some non-syntenic loci may appear in the same "linkage group", which could result in contradictions between mapping results for different mapping populations and between genetic and physical maps. Thus, up to 12% of cattle markers were assigned to wrong chromosomes and contradict the physical maps (H. Lewin, personal communication). In fish genetics,

pseudo-linkage is also a known phenomenon (see for review: [6]). To address this problem, we suggest a modified approach of clustering markers into linkage groups. In our scheme, clustering is conducted concurrently with multilocus ordering that also includes verification of the order and removing unreliable markers. Instead of using one threshold recombination rate (or LOD value), we employ a series of increasing recombination thresholds (or decreasing LODs). At the first, most stringent threshold, we have a minimum of danger of mixing markers from different chromosomes into one linkage group, but the result is a high number of linkage groups. By relaxing the stringency at the next steps, we allow end-to-end merging of the ordered linkage groups, excluding those that display the strongest affinity to each other by their interior parts.

In many practical cases, high-density mapping is associated with another difficult problem: a disproportion between a high number of scored markers and a relatively small population size. The number of markers may, by orders of magnitude, exceed the resolution of recombination for the given population size, so that only a minority of markers can be actually ordered. The question is how to choose the most informative markers to build a reliable "skeleton" map. If we consider a situation with, for example, $k \sim 1000$ markers, then for a sample size of $N \sim 100$, the minimum distance between markers that can be resolved in the map should be $\geq 1$ cM; hence, the map length for a chromosome should be $\geq 1000$ cM, which is unrealistic for the vast majority of organisms. How can the appearance of such 1000 cM maps be explained? We believe that the root is in the wrong assumption that all markers are different (resolvable by recombination). In fact, for small sample sizes, many markers comprise groups of absolutely linked markers and should be replaced by their "delegates". But even with this simplification, the number of resulting markers that differ may remain quite large, with the map length by far exceeding the expectations based on the estimates of chiasma frequencies at meiosis [7,8]. Clearly, marker scoring errors generate "false recombinants": with perfect scoring most of these recombinants would not have appeared, but after excluding absolutely linked excessive markers (replacing them by delegates), it would be possible to build an "ideal" skeletonap. Another possible complicating factor is negative interference [4,5,9,10] violating the simple principle that "the entire entity is supposed to be larger than its parts" [11].

Besides close linkage combined with a limited (small) sample size, the necessity for the selection of representative markers for the skeleton map derives from the varying information content of markers (co-dominant versus dominant, missing data, distorted segregation, and scoring errors), "absolute" linkage between repulsion-phase dominant markers, and negative interference [5]. These (or some of these) criteria are employed by other authors also. Thus, *before the analysis*, the authors also chose bin markers, whereas *after the analysis* a decision is made about excluding double recombinants and recovering missing data (usually, by assuming no recombination). The problem with the last correction aimed to reduce the map length, is that it does not affect the order of the markers. This after-ordering correction deals with maps that might have been affected by errors of marker scoring. This could cause erroneous ordering or, even worse, bringing together markers from non-homologous chromosomes.

Our objective is to get an approximation as close as possible to the true multilocus order despite the foregoing complications. A specific feature of our approach is that the choice of candidates for the skeleton map is a part of the core *ordering-verification* procedure focused on detecting and removing markers causing local map instability and non-monotonic changes of recombination (*i.e.*, deviation from the expected increase of *rf* between a marker and its subsequent neighbors). The verification process is based on multiple re-sampling runs from the scored mapping population using the so-called jackknife approach [12,13], namely, from the initial set of $N$ genotypes, we sampled a subset of $\alpha N$ genotypes (e.g., with $\alpha = 0.8$-$0.9$) scored for the same markers. The obtained sub-sample is employed to order the map. This process is applied repeatedly (e.g., 100 times), resulting in corresponding map orders. The neighborhoods that do not change upon these jackknife runs can be referred to as stable.

Clearly, such a formulation calls for massive repeated application of multilocus ordering procedures that may be computationally very challenging in case of moderate to high-density maps. Several genomic problems, including multilocus genetic mapping, in building physical maps (contig assembling for overlapping clones and radiation hybrid mappings), assembling ESTs, and others, can be formulated as *multipoint one-dimensional ordering*. Despite variation among possible optimization criteria, the one-dimensional genetic or genomic ordering problems are quite similar to the well known challenging Traveling Salesperson Problem (TSP). A powerful algorithm developed for a wide class of TSP-like Vehicle Routing Problems and referred to as the Guided Evolution Strategy (GES) [14] was successfully adapted to genetic mapping [15]. High performance and high precision of GES algorithms make them very suitable to address computation challenging multipoint ordering problems, especially in the context of our methodology requiring a verification analysis to ensure stability of the

constructed maps. The mapping of 3000 loci of a dataset generated in a maize project at the Center for Plant Genomics, Iowa State University, can be used as an example of practical efficiency of this approach [16].

## 2. METHODS, ALGORITHMS, AND EMPLOYED DATASETS

### 2.1. General Scheme

The core procedure of our approach includes the following stages (**Figure 1**):

Using marker orders rather than marker cM positions as the main map characteristic, with minimum total map length as an optimization criterion. Although map length is employed by many other authors, the stability of ordering rather than the confidence intervals of the marker positions or posterior marker positions [17,18] as a criterion of the map quality, is central to our method. To evaluate the stability of ordering, we employ re-sampling procedures [11,19].

The previously mentioned formulation was possible to implement as a mapping algorithm for many markers (*i.e.*, hundreds per chromosome) because of a novel, highly efficient method of discrete optimization developed in our lab [14]. A procedure for detecting and removing problematic markers causing local "map expansion" using the instability of local neighborhoods across re-sampling runs and the deviation from the expected monotonic change of recombination rates as criteria. A stepwise procedure of merging clusters of linked markers based on the end-to-end principle in order to reduce the danger of combining markers of non-homologous chromosomes in one map. This danger may derive from sampling deviations of corresponding recombination rates from the expected 50% or from the pseudo-linkage phenomenon [4,5,20].

### 2.2. Clustering

The first step is calculating pairwise recombination fractions ($rf$) for all pairs of markers using the maximum likelihood estimation procedure. Then, the number of clusters (linkage groups, LG) can be evaluated as a function of the threshold (maximal) value $rf_0$, allowing the preliminary assignment of a marker to a certain LG, namely, marker $m_i$ may be assigned to an $LG_j$ if recombination between $m_i$ and at least one marker from $LG_j$ is lower than the threshold $rf_0$ and is the lowest compared to the distances to any other LG. Based on the obtained information, it is necessary to choose a sufficiently small value of $rf_0$ to exclude the possibility of getting in one LG markers from non-homologous chromosomes due to pseudo-linkage (see [4,5,21,22]). But choosing an $rf_0$ that is too small will result in a large number of clusters (linkage groups) that will considerably exceed the haploid number of the species. Therefore, the next steps should include controlled merging of some of the clusters by a gradual relaxing of the conditions on pseudo-linkage (by increasing $rf$). The specific feature of our approach is that the building and ordering of the LGs are considered as interacting procedures (see **Figure 1**). If some markers of two LGs appeared closer than the relaxed $rf$, it would be reasonable to permit merging if the closest markers of the two candidate LGs are terminal or sub-terminal ("end-to-end" merging). Merging should be forbidden if the closest markers reside in the interior part of one or both candidates.

To illustrate how this scheme works, we simulated an example with two chromosomes (A and B) with pseudo-linkage. The maximum deviation from independent segregations of markers $a_i$ (chromosome A) and $b_j$ (chromosome B) was for markers with $i = 5$ and $j = 13$ (the simulated value was $rf_{a5\text{-}b13} = 0.2$, whereas the value that "occurred" was 0.19). The recombination values for consecutive adjacent markers were 0.1, excluding $r_{11\text{-}12} = 0.285$ on chromosome A and $r_{8\text{-}9} = 0.33$ on chromosome B. What happens when two different threshold values of recombination are chosen, e.g., $rf_0 = 0.3$ (a usual choice in many publications) and $rf_0 = 0.15$? With $rf_0 = 0.3$, all markers of chromosome A and the 12 last markers of

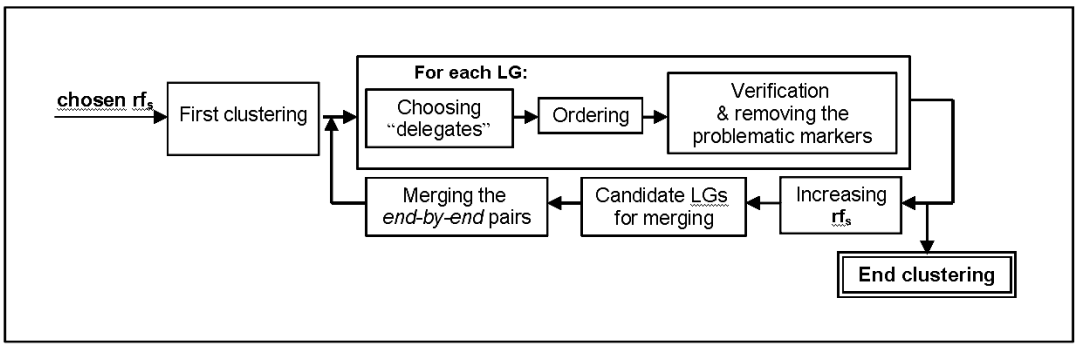



**Figure 1.** Stepwise clustering of markers in linkage groups coordinated with multilocus ordering.

chromosome B should be combined in one linkage group; the remainder of the markers of chromosome B comprised the second linkage group. The results obtained after ordering these groups followed by detecting and removing markers violating monotonic change of *rf* along the map are shown in **Figure 2(a)**. Now, we start with a more stringent threshold, $rf_0$=0.15. This choice resulted in four linkage groups (**Figure 2(b)**); upon relaxation of the threshold (0.15 → 0.30) linkage groups of the non-homologous chromosome would tend to merge, but not in the end-to-end manner (their internal parts proved to be the closest, $rf_{a5\text{-}b13}$ = 0.19). Thus, this merging was not allowed. The next step of relaxation ($rf_0$ = 0.35) with the previously mentioned rule (allowing only end-to-end merging) resulted in the correct recovery of the simulated chromosomes. The presented cycle can be repeated several times until further merging will cause an appearance of LGs with large internal gaps. Clearly, this procedure can be simplified if anchor markers are available. However, the choice and usage of anchors based on literature should be cautious because of the possibility of a relatively high level of errors in some published maps.

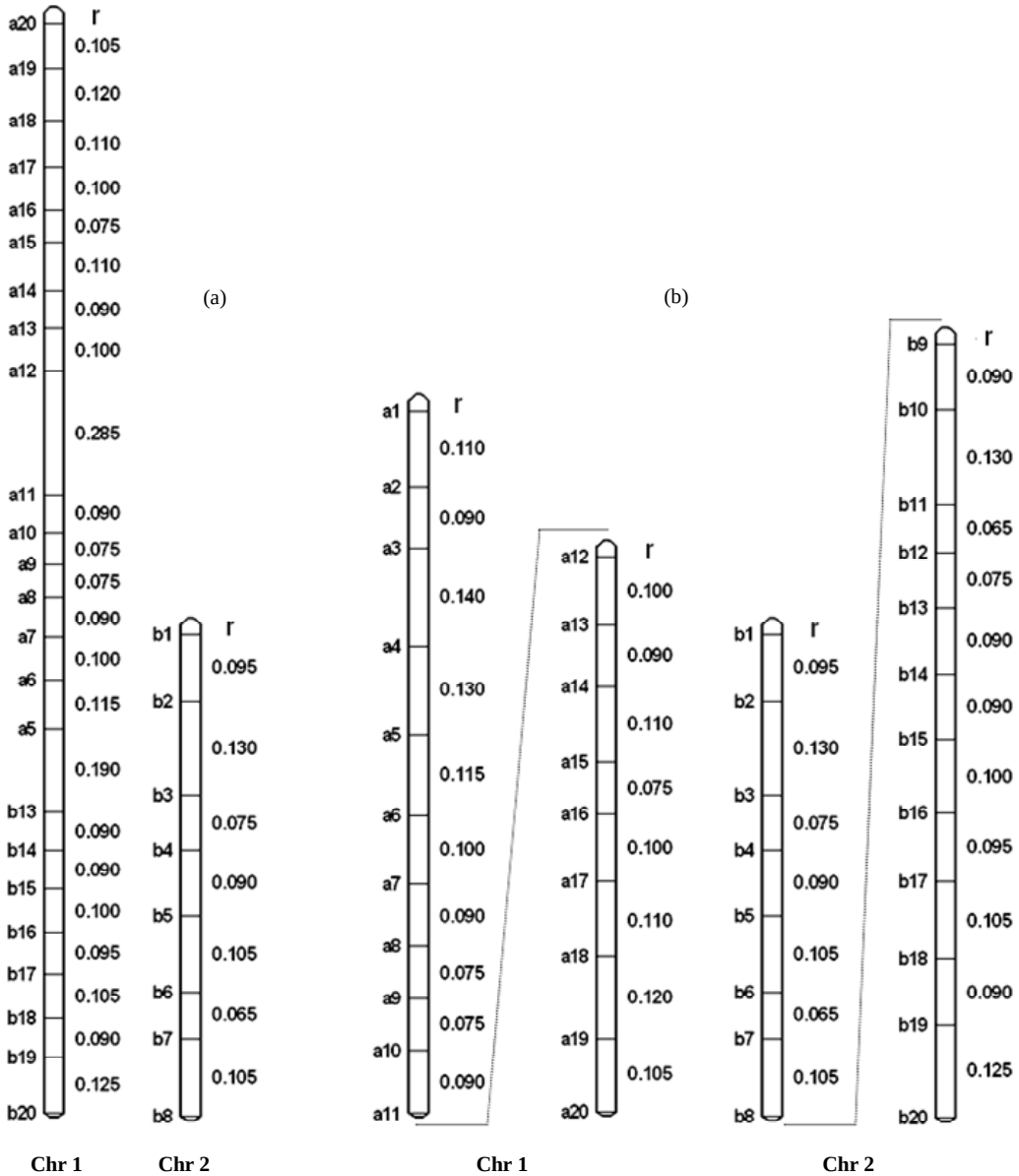



**Figure 2.** Reducing the chances of wrong clustering by stepwise relaxation of the threshold recombination: (a) Wrong clustering of markers, caused by a high threshold value of recopmbination in situations of pseudo-linkage; (b) Preventing wrong clustering by using stringent threshold recombination.

## 2.3. Multilocus Ordering

As noted above, the number of scored markers may, by orders of magnitude, exceed the number of those practically resolvable by recombination markers for the given population size. Only a small portion of markers (here referred to as *delegate markers*) can be included in the skeleton map, with the remaining markers being *attached* to the delegates. Besides the non-resolvable linkage caused by small sample size, the necessity for selection of representative markers for the skeleton map derives from non-random (clustered) recombination distribution in the genome [4], varying information content of the markers (co-dominant versus dominant, missing data, distorted segregation, and scoring errors), biased recombination estimates between repulsion-phase dominant markers [19], and negative interference [5,10]. Using our approach (implemented in MultiPoint package, http://www.multiqtl.com), one may start from a linkage group with hundreds of markers and conduct several analytical steps in order to build a reliable map:

- multilocus ordering;
- binding together closely linked markers followed by selection of delegates (bin markers) with the highest information content and replacing the groups of tightly linked markers by their "delegates"; repeated ordering and re-sampling verification of the reduced LG to detect regions of map instability;
- removing the markers causing unstable neighborhoods and violating monotonic change of recombination, followed by repeated ordering to obtain the skeleton map;
- attaching the removed markers to their best intervals on the skeleton map.

Our mapping algorithm is based on the reduction of the multilocus ordering problem to TSP or TSP-like formulation. Its main features were described earlier [11, 19]. Here we present only recent modifications and extensions of our ordering algorithm. One of the possibilities in addressing the mapping problem is to recover the marker order from a known matrix $d_{ij}$ of pair-wise marker distances based on estimates of the recombination rate. An important fact is that in genetic ordering problems the distances between the markers cannot be measured directly. For this reason, even an exact TSP solution does not guarantee that the obtained map will be robust to a small variation of the data, hence the importance of stability testing. Special formulations of the problem may include various restrictions (e.g., on a predefined order of anchor markers), implying a reduction of genetic mapping to a more complex *constrained discrete optimization* problem.

In order to improve the efficiency of our multilocus ordering algorithms, we developed a new metaheuristic approach, referred to as the Guided Evolution Strategy (GES) that combines the strengths of the Guided Local Search (GLS – [33]) and Evolution Strategies [19] in the framework of one iterative two-stage procedure. GES combines the ES and GLS metaheuristics, and these two stages are iteratively repeated until no more improvements can be found in the local search. Our experiments on 302 large-scale benchmark vehicle routing problems with constraints demonstrated that the proposed algorithm is fast, cost-effective, and highly competitive, producing the best known solutions to 82% of the constrained benchmark problems [14]. We adapted this GES approach to TSP-like problems of genetic mapping, including the Fast2Opt local search procedure and new variable (adaptive) neighborhood size. The new mapping-oriented algorithm works with small neighborhoods (25-50 neighbor markers) that allows significantly accelerate the performance on large-scale problems. The algorithm was tested on standard TSPlib problems with 50-2392 points. All known best solutions were achieved for these problems.

## 2.4. Map Verification

The objective of the verification procedure is detecting regions with unstable neighborhoods relative to the initial ordering (in the following example, explaining the method, we used 10 markers numbered from 1 to 10). This can be achieved by repeated re-sampling of the initial dataset (jackknife, bootstrap) followed by multilocus ordering for each such derivative sample (**Figure 3(a)**). Then, the identification of unstable regions can be conducted based on the frequency distribution of the right-side and left-side neighbors (**Figures 3(b)-3(d)**). The identification of such regions can be conducted by summing up corresponding neighborhood matrices (**Figures 3(b)-3(c)**) and calculating the neighborhood frequencies (**Figure 3(d)**). The higher the deviation from 1 (*i.e.*, from the "diagonal" pattern) the less certain is local order. In the example, we show only two re-sampling runs. Based on this small-size re-sampling, we can indicate certain local neighborhoods, *i.e.*, for marker pairs 1-2, 4-5, 7-8, and 9-10. In actual analysis the number of runs should be at least a few dozen or hundreds.

Clearly, the unstable neighborhoods result from fluctuations in the estimates of recombination rates across the repeated samples; the range of fluctuations depends on the sample size and the proportion of individuals taken at each jackknife run. In our framework, jackknife analysis is a modeling tool to quantify the diversity of map versions for the treated chromosome representing the sampling (stochastic) nature of the map. The results of such an evaluation can facilitate further decision

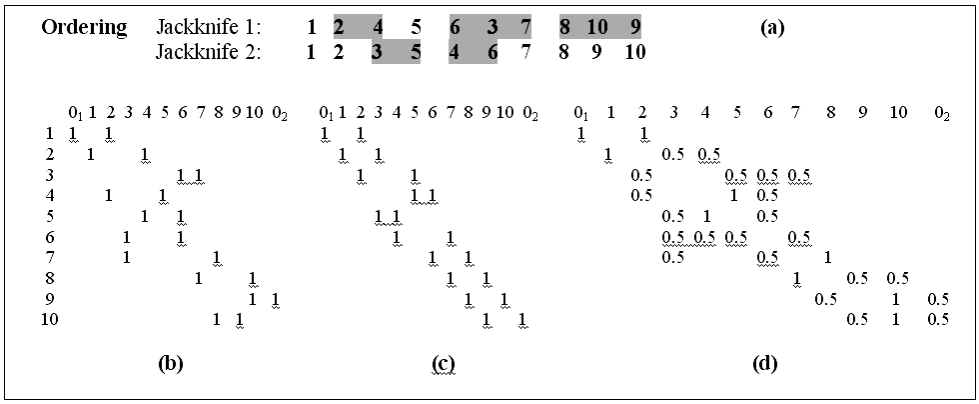



**Figure 3.** Graphical display of the verification process: (a) multilocus orders for each jackknife (the example includes only two jackknives); (b) and (c) neighborhood matrices for each jackknife run; (d) matrix of neighbourhood frequencies averaged over all runs.

making about problematic markers. These markers can be removed from the map and then the map must be re-built (**Figure 3**).

### 2.4.1. Improving Map Stability and Building Skeleton Map

After revealing the regions of map instability, we need to make a decision about the marker (or markers) responsible for local instability. For each region, there could be more than one candidate for removal from the dataset with the objective to stabilize the order. Our choice should depend on quality of the markers (anchor markers or genes vs. other markers; co-dominant vs. dominant; concordantly c-segregating with the neighbors or displaying unique segregation; fully scored or with many missing data, etc.). Taking these criteria into account, we can check the effect of removing any of the candidates using the trial-and-error approach, namely, after the removal of a candidate marker, we can re-build the map and again test its stability based on jackknife re-sampling. This computing of intensive methodology is affordable within the framework of our approach due to the very high performance of our multilocus ordering heuristic algorithm. As a result, we will come up with a stabilized (skeleton) map.

Clearly, the skeleton map will include the most reliable (informative) markers. Likewise, any group of tightly linked markers un-resolved by recombination, due to tight linkage and small sample size, will be represented by one delegate marker, which is also selected on the basis of scoring quality or biological priority (say, gene vs. anonymous marker). The three main reasons for map instability can be mentioned here: 1) tightly linked markers (with one-two recombinants in the sample) that may produce varying local orders upon jackknife runs; 2) islands of moderately linked well ordered markers separated from other neighbors by relatively large gaps that will appear in opposite orientation of the entire island (propeller effect); and 3) regions with non-mon- otonic change in recombination due to an excess of double recombinants caused by scoring errors, negative interference, or gene conversion) [5,10,20,23,24].

### 2.4.2. Criteria for Comparing Different Map Versions

To characterize the efficiency and advantages of the proposed methodology compared to other methods, we need to define some criteria used in comparisons. These include:

Map length and number of markers presented in the resulting stable map: Even with correctly unraveled marker order, the evaluated map length may be higher than the actual one, due to a certain amount of scoring errors. But deviation from the correct order will result in map length inflation even without marker scoring errors.

Controlling monotony: Some of the scoring errors may generate situations in violation of the principle that "the entire entity is supposed to be larger than its parts" [11]. Normally, for three markers ordered as *a-b-c*, one would expect: $r_{ac} > r_{ab}$ & $r_{ac} > r_{bc}$. A violation of this condition indicates that something may be wrong with the local ordering. Alternatively, a violation may be caused by negative interference or gene conversion.

Stability of local neighborhoods: With increased proportions of scoring errors, the ordering may become very sensitive even to a small sampling variation upon jackknife re-sampling. To quantify such instability, we employ a simple measure:

$$\sigma = \sqrt{(1/n)\sum_i \sigma_i^2}\ ,$$

where $\sigma^2_i$ is variance of $i^{th}$ marker neighborhood:

$$\sigma_i^2 = 0.5\sum_j p_{ij}(i-j)^2\ ,$$

and $p_{ij}$ is the proportion of jackknife runs where markers $i$ and $j$ were adjacent neighbors. The enumeration of markers is according to the multilocus order obtained on the entire sample or based on an external order (e.g., from the literature). Stable order will give $\sigma = 1$.

Concordance of segregation distortion within local neighborhoods: Segregation distortion is a known phenomenon caused by various factors [4]. Upon correct ordering, one would expect a certain correspondence between segregation ratios of neighbor markers ($seg_i$ and $seg_{i+1}$), namely, a correct order should give smaller values of the following criterion $S_i$ compared to a wrong order. If the frequency of one of the two classes at a marker locus in an RIL (dihaploid, or backcross) population or one of the homozygotes in F2 is $p_1$, then the normalized change of segregation ratio from marker $i$ to $i + 1$ can be calculated as:

$$S_i = 100 dseg_i / D_{i,j+1}$$

with $dseg_i = |p_{2i} - p_{2i+1}|$, where $p_{2i}$ and $p_{2i+1}$ are the frequencies of the second marker class for loci $i$ and $i + 1$ normalized by $p_{1i}$ and $p_{1i+1}$, and $D_{i,i+1}$ is the distance (in cM) between the markers; for F2:

$$dseg_i = \max\{|p_{2i} - p_{2i+1}|, |p_{3i} - p_{3i+1}|\}$$

where $p_{2i}$ and $p_{3i}$ are the normalized frequencies of the heterozygote and the second homozygote.

#### 2.4.3. Using Real Mapping Data for Illustration

To illustrate the efficiency of the proposed mapping strategy on real data we selected a few examples of diverse organisms: wheat, barley, oat, maize, *Arabidopsis*, mouse, rat, and trout using the data available on the web. The published results for the same datasets were employed for comparisons with our map versions.

## 3. RESULTS

An empirical assessment of the proposed analytical framework, in comparison with other procedures, can be conducted using both Monte-Carlo simulations and real data. The validation of the basic properties of our algorithms was provided in our earlier publications using simulated data [11,19]. Here we demonstrate the advantages of the extended analytical scheme using genome mapping data on several eukaryotic species: wheat, barley, oat, maize, *Arabidopsis*, mouse, rat, and trout. Obviously, our intention is just to illustrate the proposed approach rather than to revise the published maps. We believe that a revision of a map, if needed, should be conducted by the research groups generating the data.

Among the illustrations provided below, the example on wheat is presented in more detail. The results on the other examples are summarized in **Table 1**, using the case on maize to explain how we compare the published and *de novo* constructed maps. One more example, for the mouse, is summarized in a separate table (**Table 2**) due to the fact that besides chromosome 16, our results for all other chromosomes corresponded well with the published map. Like with the mouse, our map version corresponded well with the published version using a dataset on *Arabidopsis*. A slight difference was detected for chromosome 5 only due to the presence of two markers with high levels of missing data, which caused problems in the published map and were excluded from our version (see **Table 1**).

### 3.1. Wheat Chromosome 1B

Wheat data from the GrainGenes website (http://wheat.pw.usda.gov/GG2/quickquery.shtml) on chromosome 1B for the RIL mapping population Synthetic × Opata with 81 markers were employed (72 markers remained after deleting absolutely linked markers). The first step was to check the marker ordering presented on the web. Using the re-calculated pair-wise *rf* values transformed "back" to F1 level, the length of the map (with Kosambi mapping function) was estimated as $L = 444$ cM. Based on re-sampling analysis, the neighborhood stability of this map was tested and found to be relatively low (**Figure 4(a)-4(c)**).

A more stable ordering, at least for a sub-set of markers (comprising a *skeleton map*), can be achieved by removing the markers causing the observed instability and deviation from the (expected) monotonic growth of recombination rates along the map around each of the markers [11]. After such "cleaning", a skeleton map with 26 markers was obtained (**Figure 5(a)**). The first and the last markers proved to be the same, as they were in the initial map (**Figure 5(b)**); the map length was reduced from 444 cM to 138 cM (!), and this was achieved without deleting "double recombinants" and replacing missing scores by those that yield non-recombinants. The improved quality of our map was accomplished by detecting and removing problematic markers, mainly with high missing levels. Historically, these markers were placed onto the map as "second wave" and "third wave" markers that were characterized much later than the first groups of markers and for a much smaller subsample of the initial mapping population. Unlike the total set with 72 markers, the order of our skeleton map with a subset of 26 markers (see **Figure 5(a)**) corresponded well with the published map: the relative ordering of the markers in **Figure 5(b)** has only one minor "inversion" (between markers Xbcd442 (#66) and Xbcd441 (#64).

### 3.2. Maize Chromosome 1(IRIL5)

We employed data on IBM302 intercross recombinant

| Name | Missing | Segregation | Index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xpsr11(1) | 6 | 1.0:1.137 | 1 | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| XksuD14(2) | 3 | 1.0:1.383 | 2 | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| Xwhs179(3) | 13 | 1.0:1.267 | 3 | | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | |
| Xgwm550(4) | 70 | 1.0:1.250 | 4 | | | | 1 | | .3 | .7 | | | | | | | | | | | | | | | | | | | | |
| Xgwm33(5) | 53 | 1.0:1.385 | 5 | | | | | .3 | | 1 | .3 | .4 | | | | | | | | | | | | | | | | | | |
| Xgwm264(6) | 52 | 1.0:1.625 | 6 | | | | | .7 | 1 | | .3 | | | | | | | | | | | | | | | | | | | |
| Xcdo388(7) | 4 | 1.0:0.982 | 7 | | | | | | .3 | .3 | | 1 | .3 | | | .1 | | | | | | | | | | | | | | |
| Xmwg938(8) | 19 | 1.0:1.087 | 8 | | | | | | .4 | | 1 | | .6 | | | | | | | | | | | | | | | | | |
| Xmwg88(9) | 13 | 1.0:1.615 | 9 | | | | | | | | .3 | .6 | | .9 | .1 | .1 | | | | | | | | | | | | | | |
| Xabc156(10) | 22 | 1.0:1.514 | 10 | | | | | | | | | | .9 | | 1 | | .1 | | | | | | | | | | | | | |
| XGli3(11) | 4 | 1.0:1.467 | 11 | | | | | | | | | | .1 | 1 | | .9 | | | | | | | | | | | | | | |
| XksuF43a(12) | 5 | 1.0:1.444 | 12 | | | | | | | | .1 | | .1 | | .9 | | .9 | | | | | | | | | | | | | |
| XksuF43b(13) | 3 | 1.0:1.947 | 13 | | | | | | | | | | | .1 | | .9 | | 1 | | | | | | | | | | | | |
| XksuE18(14) | 13 | 1.0:2.000 | 14 | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | |
| XksuE19(15) | 9 | 1.0:1.356 | 15 | | | | | | | | | | | | | | | 1 | | .1 | | .9 | | | | | | | | |
| Xbcd340(16) | 26 | 1.0:2.423 | 16 | | | | | | | | | | | | | | | | .1 | | .3 | .1 | .7 | .6 | | | | .1 | | |
| Xrz166(17) | 6 | 1.0:1.422 | 17 | | | | | | | | | | | | | | | | | .3 | | .9 | .1 | | .1 | | | | | |
| Xbcd1124(18) | 49 | 1.0:1.357 | 18 | | | | | | | | | | | | | | | | .9 | .1 | .9 | | .1 | | | | | | | |
| Nor(19) | 62 | 1.0:1.208 | 19 | | | | | | | | | | | | | | | | | .7 | .1 | .1 | | .1 | | | | | | .8 |
| Xabg373(20) | 9 | 1.0:1.789 | 20 | | | | | | | | | | | | | | | | | .6 | | | .1 | | .1 | .1 | | .4 | | |
| Xgwm18(21) | 32 | 1.0:1.184 | 21 | | | | | | | | | | | | | | | | | | .1 | | | .1 | | .1 | | | | |
| Xcmwg645(22) | 35 | 1.0:1.353 | 22 | | | | | | | | | | | | | | | | | | | | | .1 | .1 | | .8 | | | |
| Xcdo1340(23) | 11 | 1.0:1.213 | 23 | | | | | | | | | | | | | | | | | | | | | | | .8 | | 1 | | .1 |
| Xcdo127(24) | 18 | 1.0:1.256 | 24 | | | | | | | | | | | | | | | | | .1 | | | | .4 | | | 1 | | .1 | |
| Xbcd762(25) | 3 | 1.0:1.154 | 25 | | | | | | | | | | | | | | | | | | | | | | | | | .1 | | .2 |
| Xbcd1796(26) | 8 | 1.0:1.229 | 26 | | | | | | | | | | | | | | | | | | | | .8 | | | | .1 | | .2 | |

(a)

| Name | Missing | Segregation | Index | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xbcd1796(26) | 8 | 1.0:1.229 | 26 | | .1 | | | .8 | | | | | | | | | | | | | | | | | | | | | |
| Xbcd338(27) | 1 | 1.0:1.280 | 27 | .1 | | .7 | | | .6 | | | | | | | | | | | | | | | | | | | | |
| Xcdo92(28) | 9 | 1.0:1.208 | 28 | | .7 | | .1 | | .2 | | .1 | | | | | | | | | | | | | | | | | | |
| Xbcd200(29) | 5 | 1.0:1.245 | 29 | | | .1 | | .3 | | | | | | | | | | | | | | | | | | | | | |
| Xcdo1173(32) | 6 | 1.0:1.422 | 30 | .8 | | | .3 | | .1 | | | | | | | | | | | | | .1 | | | | | | | |
| Xcdo98(33) | 4 | 1.0:1.220 | 31 | | .6 | .2 | | .1 | | .1 | .8 | | | | | | | | | | | | | | | | | | |
| Xgwm413(36) | 50 | 1.0:1.241 | 32 | | | | | | .1 | | .1 | .1 | | | | | .1 | .3 | | | | .5 | | | | | | | |
| Xmwg837(37) | 1 | 1.0:1.235 | 33 | | | .1 | | | .8 | .1 | | .2 | | | | | | | | | | | | | | | | | |
| Xbcd1449(38) | 3 | 1.0:1.383 | 34 | | | | | | | .1 | .2 | | .1 | .9 | | .6 | | | | | | | | | | | | | |
| Xcdo278(39) | 7 | 1.0:1.298 | 35 | | | | | | | | | .1 | | 1 | | | | | .9 | | | | | | | | | | |
| Xbcd12(40) | 2 | 1.0:1.404 | 36 | | | | | | | | | .9 | 1 | | .1 | | | | | | | | | | | | | | |
| Xgwm1(41) | 68 | 1.0:0.958 | 37 | | | | | | | | | | | .1 | | .1 | .4 | | | | .5 | .9 | | | | | | | |
| Xcmwg758(42) | 22 | 1.0:1.385 | 38 | | | | | | | | | .6 | | | .1 | | .1 | | | .7 | | | .1 | | | .4 | | | |
| Xbarc137(43) | 37 | 1.0:1.000 | 39 | | | | | | | .1 | | | | | .4 | .1 | | .9 | | | .1 | | | | | | | | |
| Xbarc187(44) | 35 | 1.0:1.051 | 40 | | | | | | | .3 | | | | | | | .9 | | .1 | | .4 | .3 | | | | | | | |
| Xgwm498(45) | 46 | 1.0:1.464 | 41 | | | | | | | | | | .9 | | | | | .1 | | .2 | | | .6 | | | | | | |
| Xbarc240(46) | 37 | 1.0:1.000 | 42 | | | | | | | | | | | | | .7 | | | .2 | | .8 | | | .1 | .2 | | | | |
| Xbcd386(47) | 9 | 1.0:1.255 | 43 | | | | | .1 | | | | | | | .5 | | .1 | .4 | | .8 | | .1 | | | | | | | |
| Xgwm582(48) | 57 | 1.0:1.071 | 44 | | | | | | .5 | | | | | | .9 | | | .3 | | | .1 | | .1 | | | | | | |
| Xgwm131(49) | 64 | 1.0:1.040 | 45 | | | | | | | | | | | | | .1 | | | .6 | | | .1 | | .9 | | | .3 | | |
| Xbarc181(50) | 33 | 1.0:1.216 | 46 | | | | | | | | | | | | | | | | | .1 | | | .9 | | 1 | | | | |
| Xbarc302(52) | 23 | 1.0:1.244 | 47 | | | | | | | | | | | | | | | | | .2 | | | | 1 | | .5 | | | |
| Xglk136(53) | 4 | 1.0:1.467 | 48 | | | | | | | | | | | | | .4 | | | | | | | | | .5 | | 1 | | |
| Xcdo637(54) | 6 | 1.0:1.137 | 49 | | | | | | | | | | | | | | | | | | | | .3 | | | 1 | | .7 | |
| Xbcd1150a(55) | 18 | 1.0:0.764 | 50 | | | | | | | | | | | | | | | | | | | | | | | | .7 | | .4 |
| Glu(56) | 14 | 1.0:1.061 | 51 | | | | | | | | | | | | | | | | | | | | | | | | | .4 | |

(b)

| Name | Missing | Segregation | Index | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xglk136(53) | 4 | 1.0:1.467 | 48 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| Xcdo637(54) | 6 | 1.0:1.137 | 49 | 1 | | .7 | | | | | | | | | | | | | | | | | | | | | | | |
| Xbcd1150a(55) | 18 | 1.0:0.764 | 50 | | .7 | | .4 | .2 | .7 | | | | | | | | | | | | | | | | | | | | |
| Glu(56) | 14 | 1.0:1.061 | 51 | | | .4 | | .8 | .8 | | | | | | | | | | | | | | | | | | | | |
| Xbcd1150b(57) | 5 | 1.0:1.619 | 52 | | | .2 | .8 | | .1 | | .1 | | | | | | | .8 | | | | | | | | | | | |
| Xbarc61(58) | 41 | 1.0:0.947 | 53 | | | .7 | .8 | .1 | | .1 | | | | | | | | | | | | | | | | | | | |
| Xgwm124(60) | 50 | 1.0:1.955 | 54 | | | | | | .1 | | .1 | | | | .8 | | .9 | | | .1 | | | | | | | | | |
| Xcmwg733(61) | 9 | 1.0:1.356 | 55 | | | | | .1 | | .1 | | .1 | | | | .9 | | .8 | | | | | | | | | | | |
| Xgwm153(62) | 57 | 1.0:1.900 | 56 | | | | | | | | .1 | | .4 | .9 | .5 | | | | .1 | | | | | | | | | | |
| Xgwm274(63) | 51 | 1.0:2.200 | 57 | | | | | | | | | .4 | | .6 | .1 | .2 | | | | .7 | | | | | | | | | |
| Xbcd441(64) | 2 | 1.0:1.690 | 58 | | | | | | | | | .9 | .6 | | .5 | | | | | | | | | | | | | | |
| Xmwg69(65) | 11 | 1.0:2.059 | 59 | | | | | | | .8 | | .5 | .1 | .5 | | .1 | | | | | | | | | | | | | |
| Xbcd442(66) | 7 | 1.0:1.700 | 60 | | | | | | | | .9 | | .2 | | .1 | | .1 | | .7 | | | | | | | | | | |
| Xbarc188(67) | 33 | 1.0:1.050 | 61 | | | | | | | .9 | | | | | | .1 | | .1 | .9 | | | | | | | | | | |
| Xgwm403(68) | 53 | 1.0:1.480 | 62 | | | | | .8 | | | .8 | | | | | | .1 | | .2 | .1 | | | | | | | | | |
| Xbarc81(69) | 37 | 1.0:1.516 | 63 | | | | | | | | | .1 | | | | .7 | .9 | .2 | | .1 | | | | | | | | | |
| Xbcd1562(70) | 11 | 1.0:1.476 | 64 | | | | | | | .1 | | | .7 | | | | | .1 | .1 | | 1 | | | | | | | | |
| Xbcd508(71) | 9 | 1.0:1.485 | 65 | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | |
| Xcdo1189(74) | 4 | 1.0:1.135 | 66 | | | | | | | | | | | | | | | | | | 1 | | .9 | | | | | .1 | |
| Xcdo346(75) | 4 | 1.0:0.982 | 67 | | | | | | | | | | | | | | | | | | | .9 | | 1 | | | | | .1 |
| Xgwm259(76) | 46 | 1.0:1.156 | 68 | | | | | | | | | | | | | | | | | | | | 1 | | .7 | | .3 | | |
| XksuE11(77) | 9 | 1.0:0.983 | 69 | | | | | | | | | | | | | | | | | | | | | .7 | | 1 | .3 | | |
| Xksu27(79) | 3 | 1.0:1.154 | 70 | | | | | | | | | | | | | | | | | | | | | | 1 | | .7 | .3 | |
| Xgwm140(80) | 49 | 1.0:0.941 | 71 | | | | | | | | | | | | | | | | | | | | | .3 | .3 | .7 | | .7 | |
| Xbarc80(81) | 38 | 1.0:0.974 | 72 | | | | | | | | | | | | | | | | | | | .1 | | | | .3 | .7 | | .9 |

(c)

**Figure 4.** Stability testing of wheat Synthetic × Opata standard map of 1B chromosome by using jack-knife re-sampling. The published order was chosen as a reference (diagonal). The stable order should be the one where for each marker its left-side and right-side neighbors do not vary across the repeated runs (1-1 pair along the diagonal). Notes: the numbers in brackets near the marker names indicate the marker order in the map presented on the website. (a)-(c) represent the three (overlapping) parts of the map; analysis was conducted using MultiPoint software package (http://www.multiqtl.com).

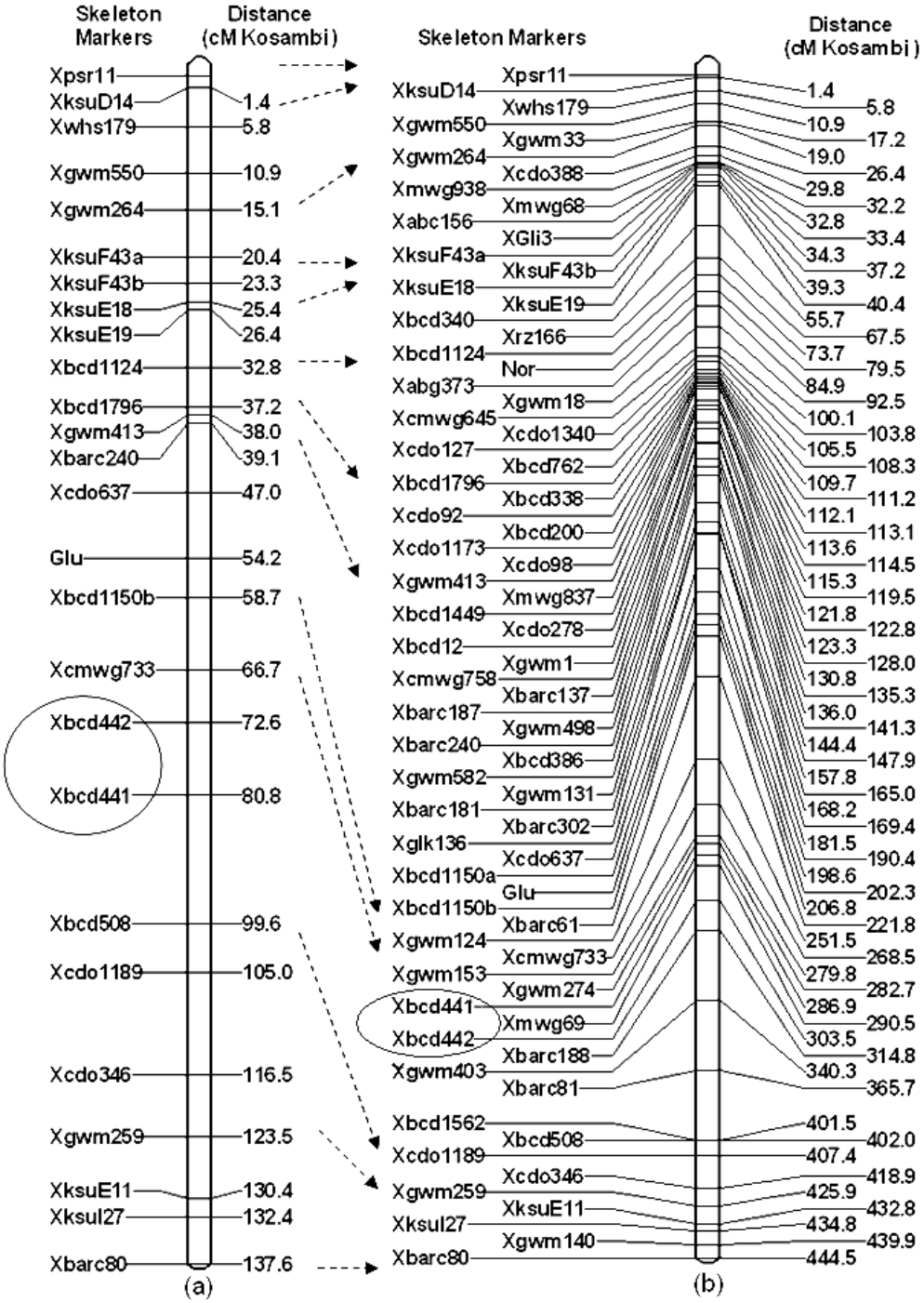



**Figure 5.** Stabilizing the map by deriving the skeleton map. The graph represents our trial to stabilize the wheat 1B map by detecting and removing the markers that caused the local instability of the neighborhoods. (a) and (b) represent the new (stabilized) and the original (web) versions, respectively. The order of skeleton markers (which proved to be the "first wave" markers) in our version of the map displays a remarkable similarity to the order on the website map, excluding one locality.

inbred line (IRIL) mapping population (http://www.maizegdb.org/cgi-bin/displaymaprecord.cgi?id=870745). To demonstrate the differences between the web version and our version of the maps the results on maize chromosome 1 are presented (**Table 1**). The length of our version of skeleton map ($L_{MP}$ = 357 cM) is half of that built with MapMaker ($L_{MM}$ = 696 cM), probably reflecting a better ordering and less discrepancy from the cy-

togenetic map (see [7,8]). Note that the distal markers in the two maps coincide. The average interval size between the neighbor markers was 696:341 = 2.05 for MM and 357:182 = 1.96 for MP.

Several examples demonstrating the problems that are typical of many published maps and are easily resolvable, via detecting and removing unreliable markers, are presented in the examples in **Table 1** for the selected organisms, including maize. Thus, disconcordant segregations of maize markers 202-203-204 in the MM map (reflected in large values of 100*dseg*/*D*) indicate that marker 203 was erroneously placed between 202 and 204. Indeed, segregation ratios for the consequent markers 202, 203, 204 in MM map are 0.78, 0.2, and 0.75,

**Table 1.** Examples of comparing the quality characteristics of the revised multilocus maps with those of the original maps.

| Species | Chr or LG | σ | | Map L, cM<br>Number of markers | | 100dseg/D | | D(ac)/D(ab or bc) |
|---|---|---|---|---|---|---|---|---|
| Oat | MN2 | MM<br>1.75 | MP<br>1.01 | MM<br>83.3<br>10 | MP<br>52<br>7 | MM<br>( 8,9)→3.4<br>(9,10)→1.6 | MP<br>(8,10)→1.2 | MM<br>D(6,8)/D(6,7) = 0.65<br>D(6,8)/D(7,8) = 0.89<br>D(8,10)/D(9,10) = 0.5 |
| Barley | Chr1H | JM<br>2.51 | MP<br>1.05 | JM<br>317<br>33 | MP<br>241<br>28 | JM<br>(17,18)→5.0<br>(18,19)→8.0 | MP<br>(17,19)→0.4 | JM<br>D(17,19)/D(17,18) = 0.67 |
| Wheat | Chr1B | JM<br>4.05 | MP<br>1.04 | JM<br>445<br>72 | MP<br>138<br>26 | JM<br>(15,16)→20<br>(16,18)→5.9 | MP<br>(15,18)→0.01 | JM<br>D(15,18)/D(15,16) = 0.55<br>D(15,18)/D(16,17) = 0.71 |
| Maize | Lg1 | MM<br>1.87 | MP<br>1.00 | MM<br>696<br>341 | MP<br>357<br>182 | MM<br>(202,203)→9.5<br>(203,204)→12.7<br>(306,307)→33<br>(307,308)→43 | MP<br>(202,204)→1.2<br>(304,308)→5.5 | MM<br>D(202,204)/D(202,203) = 0.45<br>D(202,204)/D(203,204) = 0.65<br>D(304,308)/D(304,305) = 0.28<br>D(304,308)/D(305,306) = 0.43<br>D(304,308)/D(306,307) = 0.52<br>D(304,308)/D(307,308) = 0.66 |
| Rat | Chr10 | JM<br>1.46 | MP<br>1.01 | JM<br>108<br>22(29) | MP<br>85<br>17 | JM<br>(9,10)→37<br>(10,11)→6 | MP<br>(9,11)→1.2 | JM<br>D(9,11)/D(10,11) = 0.8<br>D(23,25)/D(23,24) = 0.2<br>D(23,25)/D(24,25) = 0.4 |
| Trout | OA-1 | MM<br>1.95 | MP<br>1.06 | MM<br>136<br>25 | MP<br>88<br>16 | MM<br>(10,11)→9.4 | MP<br>(8,12)→0.5 | MM<br>D(8,12)/D(9,10) = 0.36 |
| Arabidopsis | Chr5<br>(Ler/Cvi) | JM<br>1.0 | MP<br>1.0 | JM<br>127.6<br>49 | MP<br>112.3<br>45 | JM<br>(3,4) →46<br>(4,5) →25.5<br>(20,21) →2.4<br>(21,22) →10.4 | MP<br>(3.5) →5.8<br>(20,22)→0.7 | JM<br>D(3,5)/D(4,5) = 0.52<br>D(20,22)/D(20,21) = 0.28 |

**Notes**: σ - score for neighbourhood instability (averaged for the chromosome); Map L - length of the linkage map; 100 *dseg/D* - relative score for concordance of segregation ratios (analogue of derivative for change of segregation ratios along the map); *D*(ac)/*D*(ab or bc) - ratio of the recombination distance (cM) between flanking markers of a segment to the length of one of its parts (ratio < 1 indicates wrong marker scoring, or wrong mapping of the internal marker, or strong negative interference); MM, JM, MP denote that the map was constructed with Mapmaker, Joinmap, or MultiPoint, respectively.

Coding marker names (bold denotes markers included to the skeleton map):

Oat: 6 – **p40m51n6**, 7 – bcd1230, 8 – **p48m58n4**, 9 – bcd1414, 10 – **p48m88n4**;
http://grain.jouy.inra.fr/cgi-bin/graingenes/report.cgi?class=mapdata;name=Oat%2C%20MxN%2C%20genetic%202005;show=locus;show=map;print=

Barley: 17-**E35M58-468**, 18-E37M32-209, 19-**E39M48-281**;
http://wheat.pw.usda.gov/cgi-bin/cmap/map_set_info?map_set_aid=Barley_Cebada_Capa_x_SusPtrit

Wheat: 15-**XksuE19**, 16-Xbcd340, 17-Xrz166, 18-**Xbcd1124**;
http://wheat.pw.usda.gov/GG2/quickquery.shtml

Maize: 202-**umc2237**, 203-umc2239, 204-**umc2238**, 303-**phi265454**, 304-AY110426, 305-ufg14, 306-mmp195g, 307-**npi238**;
http://www.maizegdb.org/cgi-bin/displaymaprecord.cgi?id=870745

Rat: 9-**D10Wox26**, 10-D10Mgh11, 11-**D10Wox9**, 23-**D10Wox19**, 24-D10Wox22, 25-**D10Mit7**;
http://www.well.ox.ac.uk/~bihoreau/woxtable.html#Chromosome%201

Trout: 8 – **Eacaacg176o**, 9 – Eaacctg201o, 10 – Eaagacc531o, 11 – **Eacgatc250c**, 12 – **Eaccagt121a**;
http://www.wsu.edu/%7Ethorglab/OAmap/OA2002update.xls

Arabidopsis: 3-**A.292L**, 4-nga158, 5-**BH.144L**, 20-**CH121L_Col**, 21-nga139, 22 – **DF.154.C**;
ftp://ftp.arabidopsis.org/home/tair/Maps/Ler_Cvi_RIdata

respectively. In MP map, marker 203 is removed. Removing a marker in a correct map usually results in an increased distance between the remaining markers. That was not the case in this example: $D(202,203) = 6.13$, $D(203,204) = 4.31$, and $D(202,204) = 2.80$. Thus, the presence of 203 expands the local region more than 3-fold. An even higher discrepancy in segregation ratios was detected in the neighborhood flanked by markers 304 and 308. Here three markers are problematic: 305, 306, and 307. If all five markers, from 304 till 308, are retained, the total length of the segment is equal to the sum of distances $D(304,305) = 8.2$, $D(305,306) = 5.4$, $D(306,307) = 4.4$, and $D(307,308) = 3.5$, which is **21.5**. In our version of the map markers 305, 306, and 307 are deleted, and the length of the segment becomes $D(304, 308) = 2.3$, *i.e.*, only ~1/10 of the original size of the interval flanked by 304 and 308. An additional argument in favor of our analysis is that unlike the removed markers, markers included on our skeleton map appear in the same order as in the original web map, excluding a few local discrepancies, namely, the revised local orders 2-1-4, 28-32-31-33, 47-49-48-50, 101-109-105-111, 289-295-293-292-296, and 324-331-329-334 (the numbers reflect the consequent relative positions of marker loci from 1 to 341 in the original map).

### 3.3. Mouse

We employed data on the mapping population (C57BL/6JxM.spretus)F1xC57BL/6J http://www.informatics.jax.org/searches/crossdata_form.shtml). The skeleton maps that we constructed for chromosomes 1-19 and X corresponded completely with the maps presented on the website. The only difference was for chromosome 16: the web map seems to have a serious local mistake, unless the authors used some additional information. Still, keeping in mind the high quality of the data, it may be instructive to compare our results with the original maps. For comparison, we excluded absolutely linked markers from the original maps. The results are shown in **Table 2**. For chromosome 5, the difference between our map and the web version is small and was caused by two markers (Zp3 and Ccnb1-rs1) that violated the rule that the entity cannot be smaller than its parts [11]. The two versions of the map proved identical until the marker Gusb. In the web version, the interval Gusb-Zp3 was larger than the flanking interval Gusb-Gnb2; thus, deleting Zp3 seems a reasonable suggestion. Similarly, interval D5Fcr8-Brca2 is shorter than its part Ccnb1-rs1-Brca2; thus Ccnb1-rs1 was also deleted. For chromosome 16, the considerable difference between the two versions, is caused by the erroneous placement of marker D16Fcr1 at the upper distal point of the map, despite its tight linkage with Csta.

## 4. DISCUSSION

Building correct multilocus maps is usually considered a pre-requisite for diverse genomic/genetic applications, e.g., positional cloning, anchoring contigs in physical mapping, and marker-assisted selection. Dodds and co-authors [25] surprisingly found that map errors do not seem to have too much influence on QTL mapping results. Although this may be the case in some situations, in many other situations, map errors may lead to dramatic negative impacts. Purportedly, if the objective of QTL mapping was map-based cloning, then a lot of effort might be made with no results if the map order in the region of residence of the targeted QTL was wrong. This may happen even if the assignment of the QTL to the chromosome was correct. Moreover, with some types of erroneous ordering, one could detect two QTLs in a chromosome that carried only one QTL for the considered trait, whereas under the correct ordering of markers only one QTL will be detected. Obviously, wrong ordering, even on a local scale, may also reduce the efficiency of using marker positions on the map to facilitate physical mapping.

**Table 2.** Comparing two versions of mouse maps.

| Parameter | Method | Chromosome | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 16 |
| $n_{mar}$ | MMQTX | 57 | 57 | 42 | 48 | 40 | 40 | 41 | 38 | 41 | 24 |
| | MP | 48 | 54 | 40 | 42 | 38 | 34 | 37 | 35 | 41 | 20 |
| $N_{mar}$ | | 110 | 176 | 89 | 106 | 102 | 110 | 123 | 100 | 104 | 51 |
| σ | MMQTX | 1.071 | 1.088 | 1.038 | 1.110 | 1.168 | 1.056 | 1.050 | 1.044 | 1.020 | 2.165 |
| | MP | 1.024 | 1.032 | 1.009 | 1.044 | 1.073 | 1.018 | 1.058 | 1.023 | 1.020 | 1.003 |
| $L_{chr}$, cM | MMQTX | 91.4 | 88.4 | 85.9 | 77.6 | 80.0 | 60.5 | 63.8 | 66.7 | 72.7 | 67.3 |
| | MP | 83.4 | 84.8 | 84.2 | 71.7 | 74.2 | 55.7 | 61.3 | 63.3 | 72.7 | 46.3 |

$N_{mar}$ – numebr of markers in the intial map version, $n_{mar}$ – number of delegate markers

MMQTX – Map Manager QTX

Here we propose an analytical scheme for building multilocus genetic maps that allows reasonable map quality even in complicated situations with very large number of markers, disproportionately small sample sizes, and a high level of reading errors. A high ratio of scored markers to population size has become typical in recent years due to the ever-increasing availability of high throughput genomic technologies. This new situation has encountered a psychological barrier inherited from the previous generation of the mapping community, when the number of markers was very small, justifying an intention to put each marker onto the map, even if its position is poorly resolved from the neighbor(s) or if its quality is problematic. Even now, when the number of potential markers becomes rather high (and sometimes huge), there is still a tendency to follow this approach. Some authors suggest selecting a part of the markers as *map bins* and order the vast majority of the remaining markers relative to the bins. This is a more realistic task, but still the number of markers may by far exceed the map resolution caused by small sample sizes (see Introduction section). Consequently, the resulting maps, carrying large number of markers, may be locally unreliable.

Our approach of testing map stability includes a verification procedure based on jackknife re-sampling [11]. Its major difference from the usual way of addressing this problem is that, as a measure of map stability, we consider the marker local orders [11,19] rather than the length of confidence intervals of marker map positions [17]. This approach gives flexibility in detecting and removing markers causing map instability that would be less natural to implement if the cM position on the map is the measure. Moreover, it is a well known fact that recombination rates may display very high variability between different mapping populations of the same organism, due to the effects of genotype, age, and environment [4,20]. This variability may cause serious problems in combining data from different mapping populations to build consensus maps [9,26]. If marker order is the basis of map comparisons, this problem just does not exist.

At each of the re-sampling iterations, the multilocus mapping problem is solved using TSP-like formulation. Several well known heuristic algorithms can be applied: Tabu Search, Simulated Annealing, Guided Local Search, Genetic Algorithm (with EAX crossover), Evolution Strategy, Guided Evolution Strategy, Ant Colony Behavior (ACB), and Artificial Neural Networks (ANN) (for detailed references see [15]). Currently, the most advanced software for solving unconstrained TSP is the Concorde package (http://www.tsp.gatech.edu/index.html). Concorde TSP software was applied for solving genomic problems (e.g., [27,28]. The Concorde solver uses the cutting-plane algorithm [29,30], which is an alternative to *branch and bound* to solve integer programming using the specific (one-dimensional) structure of the problem. This allows generating very good cuts helping to accelerate the optimization process, but only if the problem does not contain additional restrictions. There are examples of applying the *cutting plane* algorithm to constrained discrete optimization problems [31, 32]. In both cited studies, the authors employed a multi-processor system and parallel C++ language. In particular, using the 188 processors system, a Concorde team obtained the exact solution for a 120-point problem for 10 days. Our heuristic GES algorithm found exactly the same solution in just ten seconds using a standard PC Pentium IV (2.0 Ghz). This fact illustrates that heuristic approaches for constrained optimization problems are preferable, and GES manages with this challenge effectively. We note that our mapping-oriented GES is based on hybrid technology and employs powerful properties of both Guided Local Search [33,34] and Evolution Strategy algorithms [11,19].

Returning to the discussion about the importance of re-sampling for reliable mapping, we should stress again that one of the most frequent factors of instability in small sample sizes and in a large number of markers is errors in marker reading. As a rule, the number of such errors is rather small, just a few per marker per population. But the result is that marker pairs that had to be non-resolvable (no true recombinants) or poorly resolvable (one recombinant) upon the small sample size, become "resolvable" *and are somehow ordered*. Thus, if, on some position of the chromosome, there are several absolutely linked markers, the small rate of scoring errors, approximately equal for all markers, should disperse the markers in some multi-dimensional sphere (where they are equally distant from each other). In fact, during the mapping of these markers, they will be ordered in a one-dimensional space (as part of the map), resulting in map length inflation proportional to the number of such markers and the rate of errors. Local marker order in such a situation should be very unstable (non-reproducible) upon jackknife re-sampling, hence the proposed method of detecting such a region by our verification procedure. Removing a considerable proportion of such markers should significantly improve the map reliability and reduce the map length. As was shown in the previously mentioned example on wheat, missing data also may be an important contributor to the instability of marker neighborhoods, hence, markers with high missing levels should be considered among the first candidates for removal during the building of skeleton maps.

The discomfort that a researcher gets from a map that is too long, explains the intention of reducing the map

length, sometimes by rather artificial approaches. For example, such trials may include removing double recombinants and/or recovering the missing data by scores yielding non-recombinant genotypes after the marker ordering was finished. We found a good example for such a situation in re-analyzing the data on wheat chromosome 1B (see the previously mentioned analysis on wheat chromosome 1B). In a trial to display the map length for the multilocus order of 1B presented on the website, we used "per meiosis" recombination rates converted to Kosambi map distances. This resulted in a 1B map with L = 432 cM length, in contrast to the one published on the website with L = 104 cM. How could this huge discrepancy (432 vs. 104 cM) be explained? We managed to "reproduce" the underlying procedure with very high precision. As with many other published maps, major efforts have been invested by several teams to enrich this population with molecular markers from other populations, thereby bridging these mapping resources together. Unfortunately, new waves of markers were scored only partially, so that the level of missing data for this population was often very high, up to 50-70%.

It appeared that the problem of missing data was treated by the map constructors (or by the software they have employed) as follows. After ordering the markers, missing data were recovered by replacing the missing scores with those that resulted exclusively in non-recombinants. Clearly, the higher the level of missing data the stronger the effect of such correction will be, *i.e.*, a *reduction of the map length*. Before transforming *rf* values for RIL to *rf* values for F1, double recombinants for any pair of adjacent intervals were also replaced by non-recombinants. This method of removing erroneous double recombinants seems reasonable for F2 or double haploids, but it is inappropriate given that in RIL mapping population, "double recombinants" are not necessarily a result of scoring errors or real double recombination events. Indeed, a considerable portion of "double recombinants" have likely resulted from recombination in adjacent intervals that occurred in meiosis IN DIFFERENT generations of genotypes that remained heterozygous for those regions (in F2, F3, etc.). We considered several intervals, where our ordering was exactly the same as in the published map, but the distances were different (namely, our distances were higher than those published on the website). After conducting the previously mentioned "correction" steps, we obtained *exactly the same distances* as reported on the website map. The conducted revision analysis indicates that these distances may be irrelevant to the actual situation. Thus, the relatively small lengths of those maps are almost certainly an artifact introduced during the merging of different marker data sources, some of which contained high frequencies of missing data and inappropriate "error correction".

In genetic mapping, multilocus ordering is usually considered as a much more complicated problem compared to subdivision of the marker set into linkage groups. However, many examples indicate that markers from non-homologous chromosomes were assigned to the wrong chromosomes, presumably, due to pseudo-linkage. We have encountered these types of errors in analyzing cereal species (see [5]). Similarly, in cattle, such wrong assignments were found for 12% of markers/contigs (H. Lewin, personal communication). Thus, the pseudo-linkage phenomenon should be of special concern for large-scale genetic mapping. As indicated above, incorrect assignments can be caused by biological and statistical reasons. The probability of sampling deviation from the random segregation of markers from non-homologous chromosomes should grow with increased numbers of chromosomes, length of chromosomes, number of markers, and with small sample size. We propose in this paper that a considerable part of "wrong assignment" errors can be reduced by the algorithm of stepwise clustering markers into linkage groups alternated by multilocus ordering.

The possibility of re-sampling based on testing map stability by detecting and removing the markers that cause low map quality is the second major component of the proposed approach. Detection and removal of the markers responsible for local map instabilities and non-monotonic change in recombination rates allows building stable skeleton maps with minimal total length. Clearly, there is some degree of uncertainty in such choices; hence, there might be different versions of the skeleton map. In such situations, the high performance of our algorithms is an important advantage allowing further fast correction (complementing) of the skeleton map by using additional markers and/or some of the deleted markers. Further improvement of the mapping quality is achievable by joint analysis of mapping data from different mapping populations that can be referred to as *consensus mapping* [9,26].

## 5. ACKNOWLEDGEMENTS

The study was supported by the Israeli Ministry of Absorption and the United States-Israel Binational Agricultural Research and Development Foundation (grant # 3873-06).

## REFERENCES

[1] Lander, E.S. and Green, P. (1987) Construction of multilocus linkage maps in human. *Proceedings of the National Academy of Science*, USA**, 84(8)**, 2363-2367.

[2] Stam, P. (1993) Construction of integrated genetic link-

age maps by means of a new computer package: JoinMap. *Plant Journal*, **3(5)**, 739-744.

[3] Sapre, A.B. and Deshpande, D.S. (1987) Spontaneous emergence of parents from the F1 interspecific hybrids of Coix L. *Journal of Heredity*, **78(6)**, 357-360

[4] Korol, A.B., Preygel, I.A. and Preygel, S.I. (1994) Recombination variability and evolution. Chapman & Hall, London.

[5] Peng, J., Korol, A.B., Fahima, T., Roder M.S., Ronin, Y. I., Li, Y.C. and Nevo, E. (2000) Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: Genome-wide coverage, massive negative interference, and putative quasi-linkage. *Genome Research.* **10(10)**, 1509-1531.

[6] Korol, A.B., Shirak, A., Cnaani, A. and Hallerman, E.M. (2007) Detection and analysis of QTLs for economic traits in aquatic species. In: Liu, Z.J., Ed., *Aquaculture Genome Technologies.* Blackwell, Oxford, 169-197.

[7] Nilsson, N.O., Säll, T. and Bengtsson, B.O. (1993) Chiasma and recombination data in plants: Are they compatible? *Trends in Genetics*, **9(10)**, 344-348.

[8] Anderson, L.K., Doyle, G.G., Brigham, B., Carter, J., Hooker, K.D., Lai, A., Rice, M. and Stack, S.M. (2003) High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics*, **165(2)**, 849-865.

[9] Korol, A.B., Mester, D., Frenkel, Z. and Ronin, Y.I. (2009) Methods for genetic analysis in the *Triticeae.* Chapter 6, In: Feuillet, C. and Muehlbauer, G.J., Eds., *Genetics and Genomics of the Triticeae*. Springer, Berlin, pp. 163-199.

[10] Esch, E., and Weber, W.E. (2002) Investigation of crossover interference in barley (*Hordeum vulgare* L.) using the coefficient of coincidence. *Theoretical and Applied Genetics*, **104(5)**, 786-796.

[11] Mester, D., Ronin, Y., Minkov, D., Nevo, E. and Korol, A. (2003b) Constructing large scale genetic maps using an evolutionary strategy algorithm. *Genetics*, **165(4)**, 2269-2282.

[12] Efron, B. (1979) Bootstrap method: Another look at the jackknife. *The Annals of Statistics*, **7(1)**, 1-26.

[13] Efron, B. and Tibshirani, R.J. (1993) An introduction to the bootstrap. Chapman and Hall, New York.

[14] Mester, D. and Braysy, O. (2005) Active guided evolution strategies for large-scale vehicle routing problems with time windows. *Computers & Operation Research* **32(6)**, 1593-1614.

[15] Mester, D.I., Ronin, Y.I., Nevo, E. and Korol, A.B. (2004) Fast and high precision algorithms for optimization in large scale genomic problems. *Computational Biology and Chemistry*, **28(4)**, 281-290.

[16] Fu, Y., Wen, T.J., Ronin, Y.I., Chen, H.D., Guo, L., Mester, D.I., Yang, Y.J., Lee, M., Korol, A.B., Ashlock D. A., *et al.* (2006) Genetic dissection of maize intermated recombinant inbred lines. *Genetics*, **174(3)**, 1671-1683.

[17] Liu, B.H. (1998) Statistical genomics: Linkage, mapping, and QTL analysis. CRC Press, New York.

[18] Jansen, J., de Jong, A.G. and Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theoretical and Applied Genetics*, **102(6-7)**, 1113-1122.

[19] Mester, D.I., Ronin, Y.I., Hu, Y., Peng, J., Nevo, E. and Korol, A.B. (2003a) Efficient multipoint mapping: Making use of dominant repulsion-phase markers. *Theoretical and Applied Genetics*, **107(6)**, 1002-1112.

[20] Korol, A.B. (2001) Recombination. In: *Encyclopedia of Biodiversity.* Academic Press, San Diego, **5**, 53-71.

[21] Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L.E., *et al.* (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics*, **155(3)**, 1331-1345.

[22] Sivagnanasundaram, S., Broman, K.W., Liu, M. and Petronis, A. (2004) Quasi-linkage: A confounding factor in linkage analysis of complex diseases? *Hum Genet*, **114(6)**, 588-593.

[23] Morrell, P.L., Toleno, D.M., Lundy, K.E., and Clegg, M. T. (2006) Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. *Genetics*, **173(3)**, 1705-1723.

[24] Plagnol, V., Padhukasahasram, B., Wall, J.D., Marjoram, P. and Nordborg, M. (2006) Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. *Genetics*, **172(4)**, 2441-2448.

[25] Dodds, K.G., Ball, R., Djorovic, N. and Carson, S.D. (2004) The effect of an imprecise map on interval mapping QTLs. *Genetical Research*, **84(1)**, 47-55.

[26] Mester, D.I., Ronin, Y.I., Korostishevsky, M.A., Pikus, V. L., Glazman, A.E. and Korol, A.B. (2006) Multilocus consensus genetic maps (MCGM): Formulation, algorithms, and results. *Computational Biology and Chemistry*, **30(1)**, 12-20.

[27] Hitte, C., Lorentzen, T.D., Guyon, R., Kim, L., Cadieu, E., Parker, H.G., Quignon, P., Lowe, J.K., Gelfenbeyn, B., Andre, C., *et al.* (2003) Comparison of multimap and TSP/CONCORDE for constructing radiation hybrid maps. *Journal of Heredity*, **94(1)**, 9-13.

[28] Menotti-Raymond, M., David, V.A., Chen, Z.Q., Menotti, K.A., Sun, S., Schaffer, A.A., Agarwala, R., Tomlin, J.F., O'Brien, S.J. and Murphy, W.J. (2003) Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (*Felis catus*). *Journal of Heredity*, **94(1)**, 95-106.

[29] Dantzig, G., Fulkerson, R. and Johnson, S. (1954) Solution of a large-scale traveling salesman problem. *Operations Research*, **2**, 393-410.

[30] Applegate, D., Bixby, R., Chvatal V. and Cook, W. (1998) On the solution of traveling salesman problem. *Documenta Mathematica, Extra Volume International Congress of Mathematics III*, 645-656.

[31] Cook, W. and Rich, J.L. (1999) A parallel cutting-plane algorithm for vehicle routing problem with time window. Grant Working, Princeton University, USA.

[32] Applegate, D., Cook, W., Dash, S. and Rohe, A. (2002) Solution of min-max vehicle routing problem. *INFORMS Journal on Computing*, **14(2)**, 132-143.

[33] Voudouris, C. (1997) Guided local search for combinatorial problems. Ph.D. thesis, Department of Computer Science, University of Essex, Colchester.

[34] Tsang, E. and Voudouris, C. (1997) Fast local search and guided local search and their application to British telecom's workforce scheduling problem. *Operations Research Letters*, **20(3)**, 119-127.

**Vol.2, No.6, 590-599 (2010)**
doi:10.4236/ns.2010.26074

**Natural Science**

# Scots pine (Pinus sylvestris L.) ecosystem macronutrients budget on reclaimed mine sites—stand trees supply and stability*

**Marcin Pietrzykowski**

Department of Forest Ecology, Agricultural University of Kracow, Krakow, Poland; rlpietrz@cyf-kr.edu.pl

Received 12 February 2010; revised 24 March 2010; accepted 7 April 2010.

## ABSTRACT

**The aim of this study was to determine the sources, accumulation rate and relationships between macronutrients in reclaimed mine soils (RMS) and aboveground plant biomass on external slopes of lignite mines in central Poland. The study was conducted on two different types of sites with 10-year-old Scots (Pinus sylvestris L.) pine stands located on Quaternary loamy sands (QLS) and on Tertiary acidic carboniferous sands following neutralisation (TCS). The control plot was located in the same vicinity on an external slope in a natural pine ecosystem on a Haplic Podzol in a young mixed coniferous forest habitat (NPE). The nutrient resources, apart from N, were higher in RMS than in comparable Haplic Podzols, however, N primarily accumulated in the mineral horizons. In forest soils, the main macronutrient resources were accumulated in organic horizons, which in natural soils of coniferous forest habitats constitute the main source of nutrients. The proportion of individual macronutrients accumulated in the biomass vs. pools in soil was much lower on the external slope RMS than in the natural site, which in view of the potential richness of RMS, indicated poorer sorption and utilization of macronutrients in aboveground plant biomass than in natural habitats. Other important linear correlations (p = .05) were found between the sources of nutrients in RMS and elements accumulated in biomass (most clearly in case of K, Ca and Mg), which indicates important relationships between soil and vegetation in the first stages of ecosystem development as stimulated by reclamation.**

*This work was sponsored by the grant from Norway through the Norwegian Financial Mechanism

**Keywords:** Reclamation; Scots Pine; Ecosystem; Macronutrients

## 1. INTRODUCTION

As is true in many contemporary environmental problems, the rehabilitation of a drastically disturbed terrestrial system, such as lands mined for coal and minerals, requires site-specific knowledge to ensure the reclamation strategies chosen will be sustainable [1-16]. In central Europe, a large proportion of post-mining landscapes are reclaimed to forest.

From an ecological point of view, reclamation is a process of restoring the whole ecosystem [2]. The ecosystem, according to the traditional definition by Tensley, A.G. [17], should consist of an integrated system of biotic and abiotic elements where all the trophy layers contain a complete set of species ensuring the circulation of matter and energy flow. A complete assessment of the reclamation processes should take into consideration many ecological factors [18]. It is therefore important to determine the soil development rate including the depth of organic horizons, nutrient accumulation rates, balance of elements [3,19-22] and plant community devel- opment. Plant community parameters should include the number of species, biodiversity of communities, and the proportion of species characteristic of forest and non-forest communities [23-25].

Chemical compound pools (mineral and organic substrates) are the inanimate elements of the ecosystem (the biotope), whereas plant and animal communities are considered the animate elements of the ecosystem (biocenosis) [26]. In the course of reclamation treatment, all the factors which affect the functionality of the ecosystem are developed from essentially zero such as in primary succession [14,15]. Creation of conditions for efficient circulation of matter and energy flow between the biotope and the biocenosis determines the success of the

reclamation treatment which stimulates the process of ecosystem restoration [2]. However, the key question is when the restored biological systems cycle nutrients at rates that meet their demand without compromiseing their productivity [2,11,19,27,28].

The stand of trees is an element within animate forest ecosystems which most distinctly modifies their microclimate, light and biochemical conditions for other organisms undergoing succession. The condition and growth of stands of trees introduced to reclaimed areas directly depend on the capacity of substrates (parent rock) and the developing soil to meet the nutrient requirements which gradually increase along with the growth of biomass. There are differences between ecosystems restored on post-mining sites and natural forest ecosystems [22, 27] and they include a disturbed element circulation cycle, and low percentage of organic matter developed *in situ*. These conditions can significantly limit the pool of available elements for nutrient cycling in tree stands [29], and these cycles are also occasionally limited by direct acid-metal phytotoxicity in reclaimed mine soils (RMS).

The aim of this work was to 1) determine the sources of soil macronutrient elements: C; N; S; P; K; Ca; Mg; Na, and 2) to determine relationships between total and available forms in RMS and 3) their accumulation in the biomass of pine stands growing in the KWB 'Bełchatów' (in central Poland) reclaimed opencast lignite mine spoil heap. The study was conducted on stands of pine trees since the Scots pine (*Pinus sylvestris* L.) is one of the main species introduced when reforesting post-mining sites in central Europe [30] due to its low habitat requirements and pioneering character [31].

## 2. MATERIALS AND METHODS

### 2.1. Study Site

The study was conducted in the top portion of an opencast lignite mine spoil heap, 'Bełchatów', in central Poland (N 51 13.196; E 19 25.569). The spoil heap ranged in height from 120 to 180 m and covers an area of 1480 ha; including slopes (embankments and shelves) of 1165 ha, and a summit portion of 318 ha. Climate in the area is transitional and changeable due to clashes between polar maritime air masses and polar continental air masses. The average annual temperature is 7.6℃; the annual amplitude is 21℃; the growing period lasts 200-210 days, and total precipitation is 580-600 mm [32]. The site is located mostly on a mixture of Quaternary loamy sands, and sand with gravel, which occasionally consists of loam, bouldery clay and clay. There are also areas of Tertiary sandy strata with loam and clay, frequently carbonated and sulphurised, which are very acidic, displaying phytotoxic properties [33,34]. The reclamation treatment on the top portion of the spoil heap consisted of NPK mineral fertilisation (N − 60, P − 70 and K − 60 kg · ha$^{-1}$), and sowing a mixture of grass and leguminous plants (60 kg · ha$^{-1}$). The tertiary pyritic strata was earlier neutralised with bog lime incorporated into the surface horizon to a depth of 40 cm [34]. The area was later reforested mainly with 50% Scots pine (*Pinus sylvestris* L.) and 30% common birch (*Betula verrucosa* Ehrh.). The predominantly 1-year-old seedlings were planted on a 0.7 m length × 1.5 m width spacing [34].

### 2.2. Field Studies and Laboratory Analyses

The study plots (10 × 10 m) were located in approximately 12 to 17-year-old stands of pines on the top portion of the spoil heap (4 replications for each variant): one on the potentially fertile Quaternary loamy sand (QLS) and two on Tertiary carboniferous and pyritic sand following neutralisation (TCS). The control plot (NPE) was located in a forest in the vicinity of the spoil heap in a 17-year-old stand of pine trees, in a mixed coniferous forest habitat, ideal for this species.

Dendrometric measurements of the tree stands were made including diameter of trees at the root collar, diameter at a height of 1.3 m (*i.e.*, DBH) and their overall height (h0). Later, 35 study trees were selected proportionally to their diameter, and were cut down and weighed (the branches were weighed separately) and their foliage was tested. Foliage samples were taken to determine water content and elemental composition in the laboratory. Next, 22 tree root systems were excavated; the diameter at the root collar was measured, as were the maximum horizontal and vertical range of the roots. Roots were weighed later. Woody tissue was sampled to determine water content and elemental composition. The measurements and the obtained data were used to develop empirical equations to determine the aboveground tree biomass and root systems on the spoil heap (Mg · ha$^{-1}$) [empirical equations by Socha, 2006; unpublished data]. Fine roots less than 2 mm in diameter were assessed using cylinders (250 cm$^3$) for sampling in 3 replications at three depths (0-8 cm, 8-50 cm and 50-110 cm) from one plot per each variant on the spoil heap. In the lab, the samples were rinsed, dried and the roots were weighed. The aboveground biomass of the herbaceous vegetation was determined using the harvest method at the peak of the vegetation period, from 1 × 1 m squares in 3 replications located diagonally on each study plot. Next, herbaceous vegetation was sampled to determine water content and elemental composition in the lab.

In the course of the soil study, a 110-cm soil pit was

dug in the spoil heap and another 150 cm pit in the control plot was dug, and soil morphology was described. To determine volumetric density, samples of intact structure were collected into 250 $cm^3$ cylinders in 3 replications for each horizon. Apart from soil pits, 3 bore holes were made in each plot with soil drills [from Eijkelkamp] and mixed samples were collected to determine the content of elements and other physical and chemical soil properties from depths of: 0-8 cm (organic mineral horizons displaying some features of parent rock AiC); from 8-50 cm, and from 50-110 cm (parent rock horizons C). Samples of the organic horizon (raw humus – OL/f) were collected in autumn after the vegetation period from 1 × 1 m plots in 3 replications from each plot; the mass was assayed on the spot and mixed samples were collected for lab tests.

In the laboratory, soil samples were dried and screened with a 2 mm screen, and samples from the OL/f horizon grounded after drying. The following factors were determined in the soil samples: particle size distribution using areometrical method, pH potentiometrically in $H_20$ and in 1M KCl (1:2.5 ratio); organic carbon ($C_{org}$) and total sulphur ($S_t$) content using the infra-red absorption method, and total nitrogen ($N_t$) using the thermal conductivity method with the "Leco CNS 2000" analyser; basic exchangeable cations $Na^+$, $K^+$, $Ca^{2+}$, $Mg^{2+}$ in 1 M $NH_4Ac$ by AAS detection; the content of total elements: Na, Mg, Ca, K after digestion in the mixture of $HNO_3$, (d = 1.40) and 60% $HClO_4$ acid in 4:1 ratio using the AAS method. Phosphorus (P) in a form assimilated by plants was assayed using the Egner-Riehm method in calcium lactate extract ($(CH_3CHOHCOO)_2Ca$) acidified with hydrochloric acid to pH 3.6 and in total form using the molybdate blue colorimetry method from extracts in $HClO_4$ [35,36]. Soil subtypes were defined according to FAO taxonomy [37].

In mixed samples of Scots pine needles and herbaceous vegetation from the undergrowth (one sample for each study plot) the C, N, S content was assayed on the 'Leco CNS 2000' analyser; Na, K, Ca, Mg after digestion in the mixture of $HNO_3$ (d = 1.40) and 60% $HClO_4$ acid in a ratio of 4:1 using the AAS method and P using molybdate blue colorimetry from an extract in 60% $HClO_4$ [35].

The results, *i.e.*, the total pools of elements in soil, were statistically analyzed using the *Statistica* 6.1 programme. Differences between mean values of features from two independent groups (QLS and TCS) were tested. Distributions were compared to normality using the Shapiro-Wilks test. Next, to compare mean values of features in two variants, a t-student test was applied for independent variables ($p$ = 0.05). Correlations between sources of elements in available and exchangeable forms in soil versus accumulation in aboveground community biomass ($p$ = 0.05) were also tested.

## 3. RESULTS AND DISCUSSION

### 3.1. Soil Characteristic on the Spoil Heap

In the top portion of the spoil heap the soils were classified as Urbic Anthrosols with initial development of organic OL/f horizons which produced semi-mor-type humus at the development stage with raw humus and a thin layer of initial transitional organic-mineral horizons reflecting the features of the parent rock (AiC). In both soil types (QLS and TCS), rock type was mixed due to non-selective dumping of the rock cap. In QLS, soils developing from these strata exhibited predominantly sandy clay textures with an average of 28% silt and 4% clay. They were also sometimes interbedded with clay (43% silt fraction and 9% clay fraction) or sand. The soil bulk density averaged 1.67 g · $cm^{-3}$. In the TCS soil profile, there were remains of bog lime which had been used as a neutraliser. Soils developing on these strata exhibited lighter and more varied sandy textures, sometimes grading to loamy sand. The bulk density of the strata averaged 1.68 g · $cm^{-3}$. In the control plot in the neighboring forest ecosystem, the soil was a Haplic Podzol formed on fluvioglacial sandy strata with only up to 1% silt and up to 5% clay.

In the reclaimed areas features such as texture, soil cohesion and the neutralisation depth of toxically acidic strata determined the depth to which root systems occurred [3,38,39]. The depth to which root systems occur controls the zone of influence of living organisms and organic compounds [40]. In QLS, roots ranged to 70 cm in depth and in TCS roots penetrated to 50 cm. In those habitats, a marked flattening and deformation of pine root systems was observed. In natural conditions pines develop a typical taproot system. In the natural Haplic Podzol, the roots reached a depth of 90 cm, and has been reported that roots of pine trees in natural conditions often reach a depth of several metres [41].

The spoil heap pH in quaternary strata was neutral or alkaline and $pH_{KCl}$ averaged 7.3 and $pH_{H_2O}$ was 7.6. In organic OL/f horizons, the pH was clearly acidic (4.1 $pH_{KCl}$ and $pH_{H_2O}$ 4.4) which was due to the acidifying impact of organic litterfall under pine trees [41]. Soils on tertiary strata following neutralisation displayed different pH stratification in the soil profile. The highest pH occurred in the 0-8 cm layer of AiC horizon and averaged 5.7 $pH_{H_2O}$ and 4.9 $pH_{KCl}$. Deeper, there was a decrease in pH to as low as 3.0 $pH_{H_2O}$ and 2.7 $pH_{KCl}$. Sometimes higher content of bog lime resulted in $pH_{H_2O}$ of up to 7.8 and $pH_{KCl}$ up to 7.4. It indicated considerable mi-

cro-habitat variability in this type of plot. In natural soil, the lowest pH (4.3 $pH_{H_2O}$ and 3.5 $pH_{KCl}$) occurred in organic-mineral horizons with podzol features (AEes).

In the spoil heap, soil TEB (Total Exchangeable Bases) in QLS averaged from 26.5 to 27.6 cmol(+) · $kg^{-1}$, and CEC (Cation Exchange Capacity) from 27.0 to 28.0 cmol(+) · $kg^{-1}$ (only in organic OL/f horizon did it increase to 55.2 cmol(+) · $kg^{-1}$). The highest TEB in mineral horizons (up to 35 cmol(+) · $kg^{-1}$) was related to higher ratios of sandy clays with up to 9% clay. In TCS, TEB was much lower and ranged from 2.3 to 4.7 cmol(+) · $kg^{-1}$, whereas CEC ranged from 5.0 to 5.8 cmol(+) · $kg^{-1}$. Also in this variant, organic OL/f horizons exhibited the best exchange potential which is connected with excellent soil organic matter (SOM) exchange properties [42].

## 3.2. Community Biomass

Terrestrial ecosystems consist of above- and belowground components and their impact on one another is crucial for circulation of matter and energy flow [43,44]. Accumulation of elements in soil in the course of soil development processes, and especially the SOC sequestration potential in RMS, depends on the amount of biomass production and return to soil, and mechanisms of C protection [45].

The aboveground plant community biomass in QLS averaged 51.9 Mg · $ha^{-1}$, and in TCS it averaged 11.3 Mg · $ha^{-1}$. These differences mainly resulted from the age and stage of development of the pine trees and not from habitat conditions. However, a 19-year-old pine tree on loamy quaternary sand in QLS had a 1.5 × larger biomass than a 17-year-old pine tree in the control plot (**Table 1**).

**Table 1.** Biomass of individual components of the pine ecosystem on the top portion of 'Bełchatów' lignite mine spoil heap and in fresh mixed coniferous forest habitat.

| Variant of site | Biomass (dry biomass (Mg · $ha^{-1}$)) | | | | | |
|---|---|---|---|---|---|---|
| | Total aboveground biomass | Roots[1] | Herbaceous and shrubs | Trees | Wood[2] | Foliage |
| QLS | 51.876 (12.436) | 6.570 (1.057) | 0.134 (0.085) | 51.741 (12.454) | 43.535 (10.693) | 8.206 (1.770) |
| TCS | 11.275 (2.998) | 2.790 (0.298) | 0.033 (0.008) | 11.242 (3.005) | 9.072 (2.540) | 2.170 (0.465) |
| Control NPE | 35.813 | 7.132 | 0.152 | 35.661 | 31.863 | 3.798 |

Explanations: 394 (156) - mean (SD); [1]in natural stand (NPE) community root biomass assumed to be 0.2 of wood biomass (according to [46,47]); [2]- wood: large timber and branches of trees with DBH > 7cm.

In the forest site and in the investigated spoil heap communities, stands of trees constituted the main component of aboveground biomass while the percentage of herbaceous vegetation did not exceed 0.3%-0.4%. Compared to the total aboveground tree stands biomass, the root biomass amounted to 13% in QLS, and 25% in TCS, whereas the foliage biomass was 15% and 19% respecttively. Although the foliage comprised a small share of standing tree biomass, it made up a considerable part of annual icremental biomass production, frequently equal to that of woody tissue [48]. The typical aboveground tree biomass in forests of the temperate zone has been estimated at 21 Mg · $ha^{-1}$ (approx. 30-year-old stands of trees) and 170 Mg · $ha^{-1}$ (50-year-old stands of trees) [49]. In natural conditions on the Polish lowlands, the reported biomass of age group 1 (up to 20 year-old) stands of pine trees on average amounted to 50 Mg · $ha^{-1}$, but in the next age group it increased by nearly two-fold [50]. For four 17-year-old stands of pine trees on a reclaimed sand pit (in southern Poland) the biomass amounted to 25 Mg · $ha^{-1}$ [51]. So far, the tree biomass on the spoil heap has reached values which were close to natural conditions, and in the case of QLS, the biomass of 19-year-old stands of pine trees was higher than the control plot biomass. Very dynamic growth of the aboveground pine tree biomass on reclaimed soil was also reported in the Lusatian Mining District [8]. However, their data set refers to the first generation of stands of trees in age group 2 (*i.e.*, not exceeding 40 years of age). It is currently difficult to predict whether the cycling of nutrient elements will be intensive enough and whether a self-sustainable ecosystem will develop with such growth of aboveground tree biomass in post-mining sites [28].

## 3.3. Macronutrient Accumulation in Soil

Low content of organic matter and related low total nitrogen and organic carbon accumulation are the common limiting features of reclaimed mine soils (RMS) [19]. Soil organic matter is especially important in determining other qualities of mine soils [16,21,38,42,52]. Total acc- umulation of organic carbon ($C_{org}$) in both types of soils was similar and averaged over 54.0 Mg·$ha^{-1}$ (**Table 2**). The values were nearly 2/3-fold lower than the natural soil in the control plot where $C_{org}$ accumulation in the entire profile (up to 150 cm) exceeded 75.5 Mg · $ha^{-1}$. In organic (O) and organic-mineral (AE) horizons of the podzol developed on fluvioglacial sands under forest, $C_{org}$ accumulation ranged from 76.0 to 122.0 Mg · $ha^{-1}$. Carbon translocated to the enrichment horizon (B) should be added to this amount (the calculations have been made for Polish lowland habitat conditions on the basis of the Atlas of Polish forest soils; [53]).

**Table 2.** Pool in the soil and accumulation of elements in the biomass in pine ecosystems on the top portion of 'Bełchatów' lignite mine spoil heap.

| Element (Kg·ha$^{-1}$) | In soil | | In biomass (aboveground dry biomass) | | | |
|---|---|---|---|---|---|---|
| | Organic horizon (LO/f) | Organic-mineral and mineral horizons (AiC and C) (up to 110 cm depth) | Herbaceous and shrubs | Wood[1] | Needles | Total [2] |
| | | | Variant: QLS | | | |
| C | 1592.9 (745.0) | 52709.2 (42213.2) | 52.60 (34.88) | 23700.7 (4803.3) | 4492.0 (710.2) | 28245.3 (5497.3) |
| N | 19.1 (8.5) | 5021.2 (1247.2) | 2.61 (2.19) | 26.6 (5.4) | 87.5 (17.1) | 116.7 (21.4) |
| P | 1.8 (1.0) | 27.4 (17.6) | 0.24 (0.16) | 6.2 (1.3) | 9.5 (1.6) | 15.9 (2.8) |
| K | 3.9 (2.2) | 883.6* (269.6) | 3.76 (1.83) | 29.8 (6.0) | 49.6 (11.1) | 83.2 (16.7) |
| Mg | 2.5 (1.4) | 1288.9* (215.0) | 0.33 (0.19) | 11.0 (2.2) | 9.0 (1.2) | 20.3 (3.05) |
| Ca | 27.3 (14.2) | 94899.9* (12721.1) | 2.99 (2.23) | 42.2 (8.5) | 19.4 (2.7) | 64.6 (10.3) |
| Na | 0.1 (0.1) | 135.0* (24.0) | < 0.00 | 0.82 (0.17) | 0.25 (0.28) | 1.1 (0.42) |
| S | 1.5 (0.9) | 4890.2 (3069.5) | 0.74 (0.72) | 1.6 (0.4) | 35.3 (7.7) | 37.6 (8.5) |
| | | | Variant: TCS | | | |
| C | 1263.9 (344.6) | 53479.6 (27301.3) | 13.56 (3.11) | 5925.3 (2283.3) | 1411.2 (438.3) | 7350.1 (2717.7) |
| N | 15.0 (3.2) | 3042.5 (1226.0) | 0.29 (0.05) | 6.7 (2.6) | 31.9 (10.3) | 38.9 (12.5) |
| P | 1.8 (0.1) | 27.6 (4.7) | 0.02 (0.00) | 1.6 (0.6) | 3.2 (1.1) | 4.8 (1.7) |
| K | 2.7 (0.6) | 329.1* (45.6) | 0.44 (0.06) | 7.4 (2.9) | 20.1 (6.8) | 27.9 (9.6) |
| Mg | 2.5 (0.3) | 308.2* (140.9) | 0.03 (0.01) | 2.8 (1.1) | 2.4 (0.8) | 5.2 (1.9) |
| Ca | 29.6 (7.4) | 9778.1* (3215.4) | 0.16 (0.08) | 10.5 (4.1) | 9.4 (3.5) | 20.1 (7.5) |
| Na | 0.12 (0.03) | 95.04* (8.90) | < 0.00 | 0.2 (0.1) | 0.04 (0.03) | 0.24 (0.10) |
| S | 1.3 (0.2) | 4914.0 (2376.6) | 0.02 (0.01) | 0.3 (0.1) | 10.8 (5.0) | 11.1 (5.1) |

Explanations: 19.1 (8.5) - mean (SD); n = 4 (number of plots in variant); [1]wood biomass of trees with DBH > 7 cm; [2]total element's accumulation in aboveground biomass (trees, herbaceous and shrubs biomass), *differences for soil element's resource are significant at $p$ = 0.05 level (T-student test).

Nitrogen (and phosphorus) is one of the most deficient elements in reclaimed sites [5,54]. In the spoil heap soils of former lignite mines developing on tertiary carboniferous sands, geological carbon in the form of lignites often occurs and complicates soil C analysis. Therefore, the C and N accumulation rate in the deeper horizons of the spoil heap soils, especially in case of TCS, may actually be overestimated. Although total N accumulation increases in the course of soil development in mine soils, both in those undergoing reclamation treatment and those where natural succession is taking place [5,12, 14,38], it has been found that the average annual accumulation of N fluctuates and may change with the age of soil and vary by community type introduced in reclamation seedings [38].

Moreover, N accumulation is much less dynamic than carbon accumulation [14]. Also, in soils where primary succession takes place, N is gradually mineralised [54] which may be deficient for rapidly growing young trees in reclaimed areas which require a lot of N over time [8]. Total accumulation of $N_t$ in QLS soils was on average 5.0 Mg · ha$^{-1}$, and therefore 2/3-fold higher than in TCS soils where it amounted to 3.0 Mg · ha$^{-1}$. In comparable natural soils, N accumulation was much higher at 11.0 Mg · ha$^{-1}$.

The highest accumulation of exchangeable $Ca^{2+}$, $Mg^{2+}$, $K^+$, $Na^+$, and available P in mineral horizons took place in QLS on Quaternary strata, and in all cases (except for sulphur) it was considerably higher ($p$ = 0.05) than in TCS (**Table 2**). High sulphur accumulation ($S_t$), reaching 4.9 Mg · ha$^{-1}$, was related to the properties and origin of spoil heap strata. Soils developing on Tertiary carboniferous and pyritic strata may contain more than 1% sulphur and are referred to as "sulphurous mine soils"

[33]. The nutrient resources accumulated in mineral horizons of a comparable natural podzol were much lower than in QLS with $Ca^{2+}$ = 132; $Mg^{2+}$ = 42; $K^{+}$ = 16; P = 1.4; and $Na^{+}$ = 2.5-fold lower, respectively, than in TCS with 13; 10; 6; 1.3; and 1.7-fold respectively. According to a habitat classification based on the Soil Quality Index used in forestry in Poland [53], QLS soils could be even classified higher than natural deciduous forest sites (eutrophic). The accumulation ratio of individual elements in mineral horizons (up to 110 cm depth) and in organic horizons - raw humus layer OL/f in RMS and Ol + Olf in Haplic Podzol ($MH_{BA}$:$OH_{BA}$) (**Table 3**) constituted a major difference in the distribution of elements accumulated in RMS on the spoil heap and in the comparable Haplic Podzol. In initial mine soils, the OL/f horizons were still insufficiently developed and did not have a sufficient pool of macronutrients.

Furthermore, SOM accumulation, decomposition, and mineralization were probably not well established enough to meet the nutrient supply needs for vegetation as in natural forest habitats [29,55]. In the oligotrophic Haplic Podzol, the resources of elements in mineral horizons were the same (in case of C and Na), and nearly the same (Ca, Mg and K) or even lower (P), and only in exceptional cases such as nitrogen ($MH_{BA}$:$OH_{BA}$ accumulation was 4.6) and sulphur ($MH_{BA}$:$OH_{BA}$ was 558) was it many times higher in comparison to resources in the organic horizons (**Table 3**).

**Table 3.** The ratio of the accumulation of macronutrient resources in soil mineral horizons to accumulation in organic soil horizon ($MH_{BA}$:$OH_{BA}$) and of biomass to soils ($B_{BA}$: $S_{BA}$) on the spoil heap of KWB 'Bełchatów' in Quaternary loamy sand strata (QLS variant) and in Tertiary carboniferous and pyritic sands following neutralisation (TCS variant) and in a natural pine ecosystem on Haplic Podzol in fresh mixed coniferous forest habitat (NPE).

| Variant of site/ element | $MH_{BA}$:$OH_{BA}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | C | Ca | Mg | K | P | S | Na | N |
| QLS | 33.09 | 3471.73 | 515.56 | 226.60 | 15.54 | 3343.72 | 964.04 | 263.03 |
| TCS | 42.34 | 330.26 | 192.30 | 120.88 | 19.33 | 3780.00 | 826.40 | 203.14 |
| NPE | 0.99 | 1.24 | 0.64 | 0.62 | 0.34 | 558.20 | 0.96 | 4.64 |
| | $B_{BA}$:$S_{BA}$ | | | | | | | |
| QLS | 0.54 | 0.001 | 0.02 | 0.09 | 0.58 | 0.008 | 0.008 | 0.023 |
| TCS | 0.14 | 0.002 | 0.02 | 0.08 | 0.19 | 0.002 | 0.003 | 0.013 |
| NPE | 0.63 | 0.45 | 15.42 | 4.70 | 4.05 | 0.023 | 0.072 | 0.266 |

Explanation: values calculated based on mean for variants; MH – soil mineral horizons up to 110 cm depth; BA – element accumulation or source in ($Mg \cdot ha^{-1}$); OH – organic horizons (row humus layer OL/f in reclaimed soil and Ol + Olf in Haplic Podzol); B – aboveground biomass; S – soil; variant's abbreviation and element's form in soil - see methodology chapter

## 3.4. Relationships between Elements in Soil

The (C:N ratio) may be regarded as an indicator of changes in soils including intensification of organic matter mineralisation processes and related N-availability to plants during the decomposition of organic matter in soil [29]. In the investigated spoil heap, the soil C:N ratio in the OL/f horizon exceeded 80, whereas in QLS variant in the AiC horizon was 11 and in TCS variant it was 16. In the control podzol the C:N ratio in organic horizons was lower and was 51, and in the organic-mineral horizon (AEes) it was 17. For mor type humus characteristic of podzols in temperate climatic zones, the C:N ratio in organic and organic-mineral horizons oscillates between 30 and 40 and sometimes reached higher values [29]. Scots pine as a species characteristic of coniferous forests produces organic litterfall which decomposes with difficulty and the C:N ratio usually exceeded 70 [41]. It was assumed that for initial soils on post-mining sites, the C:N ratio in organic-mineral horizons below 25 would indicate regular mineralization processing of organic matter [56].

The potential of the developing mine soils to meet plant nutrient requirements depended on the percentage of elements in forms available for plants (for this study: $Na^{+}$, $K^{+}$, $Ca^{2+}$, $Mg^{2+}$ in exchangeable form and available P). In the organic horizons, these forms depended directly on the decomposition rate and mineralization of organic matter developed *in situ*. In mineral horizons, they largely depended on the weathering rate of minerals in the substrate. In natural habitats and especially in oligotrophic podzols, nutrients were mainly stored in the organic horizons and they were gradually released via mineralization processes [29]. This was of key importance in providing nutrients to trees as a limited amount of nutrients in soil could be compensated for by quick biological cycling of elements [55]. Soil organic matter (SOM), even though in its initial phase of accumulation, plays an important role in the tree nutrition balance in reclaimed areas [42,52,57]. In the investigated podzol, the highest percentage of exchangeable $Ca^{2+}$, $Mg^{2+}$, $Na^{+}$ and P (in available form) compared to total forms occurred in organic horizons (Ol and Ofh; 0-9 cm) and amounted to 43, 49, 35 and 18%, respectively of the total elemental pool. In the case of $K^{+}$, the highest percentage of exchangeable forms in the total pool of elements occurred in the enrichment horizon (Bfe, 50-94 cm) and amounted to 6% (**Figure 1**). In RMS (QLS), the most favourable relationship in this respect also occurred in organic horizons OL/f (0-2 cm), where $Mg^{2+}$ amounted to 41% (of total Mg) whereas $Ca^{2+}$ was 24%; $K^{+}$ = 43%; $Na^{+}$ = 24% and P was = 19% of the total elemental pool (**Figure 1**).

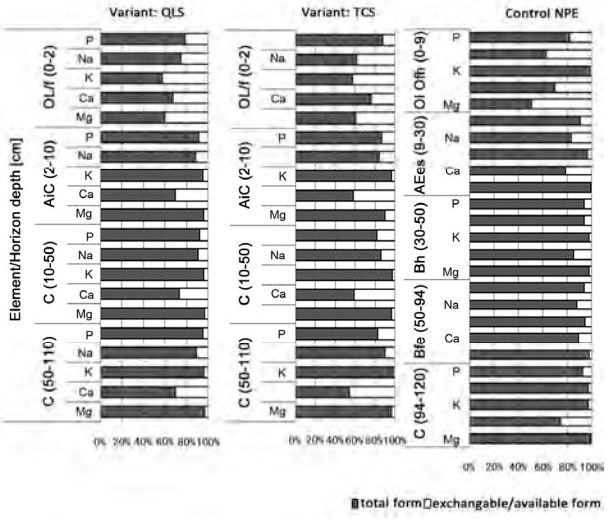



**Figure 1.** The share of exchangeable and available forms of the total macronutrient elemental pool in spoil heap soils of KWB 'Bełchatów' in Quaternary loamy sand strata (QLS variant) and in Tertiary carboniferous and pyritic sands following neutralisation (TCS variant) and in a natural pine ecosystem on Haplic Podzol in fresh mixed coniferous forest habitat (NPE).

In mineral horizons, the percentage of exchangeable and available forms decreased. However, in comparison to natural soils, the percentage of exchangeable forms in the entire pool of macronutrient elements was considerably higher. In the deeper mineral horizons, nutrients were mainly supplied due to weathering of minerals. In the more shallow horizons, what was also important was the enrichment process occurring via soil development where organic matter, colloidal fractions and sesquioxides were being supplied from the organic layers [58]. In QLS soils in the OL/f horizon (0-2 cm), similar relationships were noted for $Mg^{2+}$, $K^{+}$, and $Na^{+}$. Exchangeable forms of these elements constituted approximately 40 % of the total elemental pool in total form. However in the case of $Ca^{2+}$ and P, the largest percentage of forms available to plants occurred in mineral layers: P at 17% from 10 to 50 cm, and $Ca^{2+}$ at 46% from 50 to 110 (**Figure 1**). This was clearly connected with the bog lime neutralization treatment which was incorporated to a depth of at least 30 cm.

### 3.5. Relationships between Element AccuMulation in Soils and Aboveground Biomass

Relations between soils and vegetation under natural conditions have long been studied and soil development and ecological succession of communities are closely linked [26,47,59]. In reclaimed post-mining sites these relationships are not yet stable and may be frequently disturbed. Clear links between the trophism of mine soils (as expressed by Soil Trophy Index (according to Brożek and Zwydak [53]), and ecological indicators based on Ecological Indicator Values of Vascular Plants (according to Ellenberg, [60]) have been documented in sand pits where natural succession was allowed to occur ('Szczakowa' sand mine pit in southern w Poland; Pietrzykowski and Krzaklewski [61]). Relationships between community features and abundance of soils developing on former mining sites under succession were also described by Wali [14] on the bases of studies of abandoned coal-mine spoil materials in a mixed grass prairie region (in western North Dakota, USA).

For this spoil heap, the ratio of elements accumulated in the aboveground biomass to the resources in soil mineral horizons ($B_{BA}$:$S_{BA}$) differed largely from the control plot in the mixed coniferous forest, especially for Ca, Mg, K and P. Communities in reclaimed areas accumulated considerably less of those elements in relation to the potential resources (**Table 3**). For biomass in the forest habitat, the $B_{BA}$:$S_{BA}$ ratio was much higher, which indicates that the elements available in soil were much better utilized by those communities. On this basis, it may also be claimed that the element exchange mechanism by pine communities on the spoil heap has different dynamics than in natural habitats. A dependence analysis between the accumulation of nutrients (in exchangeable and available forms expressed in Mg · $ha^{-1}$ to a depth of 110 cm) and elements accumulated in community biomass showed a significant linear correlation ($p$ = 0.05) for K, Ca and Mg (**Table 4**). This indicates the existence of marked relationships between the abundance of soil nutrients available to plants and the level of elements accumulated in community biomass developing on the spoil heap. Under natural conditions, such obvious dependence occurs in the first stages of primary succession where plant communities depend directly upon elements from the parent rock transformed into soil [26,49]. In the more complex conditions of natural forest ecosystems there are many other variables which modify these factors, including organic matter decomposition rate, individual biochemical cycles of elements, and the soil volume used by tree root systems which was difficult to determine in this study.

## 4. CONCLUSIONS

The highest aboveground pine stand biomass occurred in QLS on quaternary strata, However when compared to TCS on less abundant tertiary sands following neutralization, lower biomass resulted from age differences (19 and 12 years), and not just from differences in habitat

**Table 4.** A table of correlations between resources of macronutrient elements (Mg · ha$^{-1}$) in mineral soil horizons to 110 cm depth versus those in aboveground pine tree stand biomass on the upper portion of "Bełchatów" lignite mine spoil heap.

| In Soil (mineral horizon up to 110 cm depth) | Element's accumulation (Mg · ha$^{-1}$) | In biomass (aboveground biomass: trees and herbaceous vegetation) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | C-biom | N-biom | S-biom | P-biom | K-biom | Ca-biom | Mg-biom | Na-biom |
| | C-soil | 0.18 | 0.11 | 0.30 | 0.17 | 0.31 | 0.22 | 0.11 | 0.09 |
| | $N_t$-soil | 0.55 | 0.59 | 0.55 | 0.57 | 0.52 | 0.58 | 0.61 | 0.44 |
| | $S_t$-soil | 0.03 | -0.01 | -0.10 | 0.05 | 0.02 | 0.01 | 0.02 | -0.06 |
| | P-soil | -0.05 | 0.04 | -0.17 | -0.03 | -0.18 | -0.07 | 0.03 | 0.06 |
| | $K^+$-soil | **0.80*** | **0.74** | **0.84** | **0.78** | **0.85** | 0.82 | 0.78 | 0.61 |
| | $Ca^{2+}$-soil | **0.92** | **0.90** | **0.91** | **0.92** | 0.92 | **0.93** | 0.93 | **0.79** |
| | $Mg^{2+}$-soil | **0.82** | **0.80** | **0.81** | **0.82** | 0.82 | 0.83 | **0.85** | 0.67 |
| | $Na^+$-soil | 0.62 | 0.56 | 0.59 | 0.61 | 0.64 | 0.62 | 0.62 | 0.37 |

*__marked__ differences are significant at $p = 0.05$; n =8; $N_t$ - total nitrogen; $S_t$ - total sulphar; $K^+$ - exchangeable cation forms – see methodology chapter.

conditions between the plots. The higher community biomass in QLS (2/3-fold) on the spoil heap compared to pine tree stand biomass in the control plot in natural habitat indicated that quaternary loamy sand strata was potentially a good soil substrate for this species. Total accumulation of organic carbon ($C_{org}$) in both soil variants was 2/3-fold lower, and in case of total $N_t$, more than 3 -fold lower when compared to natural soil in the control plot. Nitrogen was the most deficient element in those conditions. The accumulation and the biogeochemical cycling of carbon and nitrogen were closely linked with the processes of soil development and community development on the spoil heap. The accumulation of these elements was a good indicator of the rate of these processes. In case of the exchangeable cations ($Ca^{2+}$, $Mg^{2+}$, $K^+$, $Na^+$) in the mineral RMS horizons, their accumulation was mostly connected with the potential abundance of rock strata and weathering processes during the development of soil. This was why the resources of these elements were considerably higher in the spoil heap soil than in the Haplic Podzol which formed on oligotrophic fluvioglacial sands. The higher (statistically significant) resources of macronutrients in QLS soils on quaternary strata compared to TCS soils on tertiary sands were also related to the origin and properties of parent rocks of the developing soil. Since this feature clearly differentiated the degrees of soil nutrient abundance on the spoil heap, it may be used to develop a habitat condition indicator for these materials.

A significant difference between RMS soils and natural Haplic Podzols were the accumulation ratios of individual elements in mineral and organic horizons ($MH_{BA}$: $OH_{BA}$). In the case of mine soils, the initial organic horizons did not yet constitute a significant source of nutrients and SOM accumulation and decomposition were not the basic mechanism for supplying plants with nutrients as is the case in natural forest habitats. There were also differences in the ratio of elements accumulated in aboveground biomass to the potential sources in soil ($B_{BA}$:$S_{BA}$) on the spoil heap and in the control plot, particularly for Ca, Mg, K and P. Plant communities in reclaimed area accumulated much fewer elements compared to potential sources in soil. However, for plant biomass in forest habitats of the oligotrophic Haplic Podzol, the ratio was much higher and indicated that macronutrient resources in soil were optimally utilized by the plant community. On this basis it may be assumed that the exchange mechanism of elements by plant communities dominated with pine on the spoil heap had different dynamics than in natural habitats. Moreover, the reported correlations between the accumulation of nutrients in soil and elements accumulated in plant community biomass (most clearly in the case of K, Ca and Mg), indicates the existence of marked links between soil and vegetation in the process of ecosystem development on a former mining spoil heap as stimulated by reclamation treatments.

## 5. ACKNOWLEDGEMENTS

This study was financially supported by a grant from Norway through the Norwegian Financial Mechanism. The author also thanks Professor W. Lee Daniels Ph.D. from Virginia Tech for critical text correction; Jarosław Socha Ph.D. from Department of Forest Mensuration for his kind assistance in the process of preparing statistical analysis, and laboratory staff of Department of Forest Ecology and Department of Forest Soil Science ACU Krakow, Poland.

## REFERENCES

[1] Bradshow, A.D. and Chadwick, M.J. (1980) The restoration of Land. The Ecology and reclamation of derelict and degraded land. Blackwell Scientific Publications, Oxford.

[2] Bradshaw, A.D. and Hüttl, R.F. (2001) Future minesite restoration involves a broader approach. *Ecolgical Engi-*

*neering*, **17(2-3)**, 87-90

[3] Daniels, W.L., Genthner, M.H. and Hodges, R.L. (1992) Soil development in sandy tailings derived form mineral sands mining in Florida. *Proceedings of the* 9*th National Meeting of the American Society for Surface Mining and Reclamation*, Duluth, 14-18 June 1992, ASSMR, Lexington, 37-47.

[4] Daniels, W.L., Alley, M.M., Zelazny, L.W. and Schroeder, P. (1996) Strategies for rehabilitating prime farmland following mineral sands mining in Virginia. *Annual Research Report, Submitted to RGC* (*USA*) *Mineral Sands Inc*, 1-20.

[5] Daniels, W.L., Schroeder, P.D., Nagle, S.M., Zelazny, L.W. and Alley, M.M. (1999) Reclamation of prime farmland following mineral sands mining in Virginia. *Proceedings of the* 16*th National Meeting of the American Society for Surface Mining and Reclamation*, "*Mining and reclamation for the Next Millennium*", Scottsdale, Arizona, 13-19 August 1999, ASSMR, Lexington, 146-156.

[6] Daniels, W.L., Evanylo, G.K., Nagle, S.M. and Schmidt, J.M. (2001) Effects of biosolids loading rate and sawdust additions on row crop yield and nitrate leaching potentials in Virginia sand and gravel mine reclamation and Reclamation—A different approach. *Proceedings of the* 18*th National Meeting of the American Society for Surface Mining and Reclamation*, Albuquerque, 3-7 June 2001, ASSMR, Lexington, **2**, 399-406.

[7] Daniels, W.L., Nagle, S., Whittecar, G.R. and Evenylo, G. (2002) Effect of Biosolides application on ground water nitrate-N levels in sand and gravel mine reclamation in Virginia. *Proceedings of the* 19*th National Meeting of the American Society for Surface Mining and Reclamation*, Lexington, 9-13 June 2002, ASMR, 99-113.

[8] Hüttl, R.F. and Weber, E. (2001) Forest ecosystem development in post-mining landscapes, a case study of the Lusatian lignite district. *Naturwissenschaften*, **88(8)**, 322-329

[9] Krzaklewski, W. (1982) Die Möglichkeiten der ausnutzung der natürlichen vegetation in der rekultivierung der Nachabbaugelände (Abraum-oder Bergehalden). *Kommunalverband Ruhrgebit Conference*, Essen, 7-10 September 1982.

[10] Pietrzykowski, M. (2006) Properties of soils formed on reclaimed areas and areas left for succession on post-exploitation sand filing excavation (in Polish, English summary). *Soil Science Annual*, **57(3-4)**, 97-105.

[11] Pietrzykowski, M. and Krzaklewski, W. (2007a) An assessment of energy efficiency in reclamation to forest. *Ecological Engineering*, **30(4)**, 341-348

[12] Pietrzykowski, M. and Krzaklewski, W. (2007b) Soil organic matter, C and N accumulation during natural succession and reclamation in an opencast sand quarry (southern Poland). *Archives of Agronomy and Soil Science*, **53(5)**, 473-483

[13] Urbanska, K.M., Webb, N.R. and Edwards, P.J. (Eds.) (1997) Restoration Ecology and Sustainable development. Cambridge University Press, New York.

[14] Wali, M.K. (1999) Ecological succession and the rehabilitation of disturbed terrestrial ecosystems. *Plant and Soil*, **213(1-2)**, 195-220.

[15] Wali, M.K. and Freeman, P.G. (1973) Ecology of some mined areas in North Dakota. in: Wali, M.K., Eds., *Some Environmental Aspects of Strip Mining in North Dakota*. Education Series 5, North Dakota Geological Survey, Grand Froks, 25-47.

[16] West, T.O. and Wali, M.K. (2002) Modeling regional carbon dynamics and soil erosion in disturbed and rehabilitated ecosystems as affected by land use and climate. *Water, Air and Soil Pollution*, **138(1-4)**, 41-163.

[17] Golley, F.B. (1993) History of the Ecosystem Concept in Ecology. Yale University Press, New Haven.

[18] Rodrigue, J.A., Burger, J.A. and Oderwald, R.G. (2002) Forest productivity and commercial value of pre-law reclaimed mined land in the eastern United States. *Northern Journal of Applied Forestry*, **19(3)**, 106-114

[19] Bendtfeldt, E.S., Burger, J.A. and Daniels, W.L. (2001) Quality of amended mine soils after sixteen years. *Soil Science Society of America Journal*, **65**, 1736-1744

[20] Bradshaw, A.D. (1983) The reconstruction of ecosystems. *Journal of Applied Ecology*, **20**, 1-17

[21] Li, R.S. and W.L. Daniels (1994) Nitrogen accumulation and form over time in young mine soils. *Journal of Environmental Quality*, **23**, 166-172

[22] Schaaf, W. (2001) What can element budgets of false-time series tell us about ecosystem development on post-lignite mining sites? *Ecological Engineering*, **17(2-3)**, 241-252

[23] Jochimsen, M.E.A. (1996) Reclamation of colliery mine spoil founded on natural succession. *Water Air Soil Pollution*, **91(1-2)**, 99-108

[24] Pietch, W.H.O. (1996) Recolonization and development of vegetation on mine spoils following brown coal mining in Lusatia. *Water Air Soil Pollution*, **91(1-2)**, 1-15.

[25] Wiegleb, G. and Felinks, B. (2001) Primary succession in post-mining landscapes of lower Lusatia—Chance or necessity. *Ecological Engineering*, **17(2-3)**, 199-217.

[26] Odum, E. (1971) Fundamentals of ecology. 3rd Edition, W. B. Saunders Co., Philadelphia.

[27] Bell, L.C. (2001) Establishment of native ecosystems after mining—Australian experiences across diverse biogeographic zones. *Ecological Engineering*, **17(2-3)**, 179-189

[28] Knoche, D., Embacher A. and Katzur, J. (2002) Water and element fluxes of red oak ecosystems during stand development on post-mining sites (Lusatian Lignite District). *Water, Air, and Soil Pollution*, **141(1-4)**, 219-231

[29] Baule, H. and Fricker, C. (1970) The fertilizer treatment of forest trees. BLV-Verlagsges, Munchen.

[30] Baumann, K., Rumpelt, A., Schneider, B.U., Marschner, P. and Hüttl, R.F. (2006) Seedling biomass and element content of Pinus sylvestris and Pinus nigra grown in sandy substrates with lignite. *Geoderma*, **136**, 573-578

[31] Fober, H. (1993) Żywienie mineralne. In: Białobok, S., Boratyński, A. and Bugała, W., Eds., *Biologia sosny zwyczajnej*, PAN Instytut Dendrologii, Sorous Press, Poznań-Kórnik, 182-193

[32] Woś, A. (1999) Climate of Poland (Klimat Polski) (in Polish). PWN Press, Warsaw.

[33] Katzur, J. and Haubold-Rosar, M. (1996) Amelioration and reforestation of sulfurous mine soils in Lusatia (Eastern Germany). *Water, Air and Soil Pollution*, **91(1-2)**, 17-32

[34] Krzaklewski, W., Kowalik, S. and Wójcik, J. (1997) Re-

clamation of acidic and toxic strata in brown coal opencast mining. MONOS Press, Krakow.
[35] Ostrowska, S., Gawlinski, Z. and Szczubialka, Z. (1991) Procedures for soil and plants analysis (in Polish). Institute of Environmental Protection, Warsaw.
[36] Van Reeuwijk, L.P. (1995) Procedures for soil analysis. 5th Edition, Technical Paper 9, ISRIC, FAO, Wageningen.
[37] FAO-UNESCO (1997) Soil map of the World—revised legend. ISRIC, Wageningen.
[38] Anderson, D.W. (1977) Early stages of soil formation of glacial till mine spoils in semiarid climate. *Geoderma*, **19(1)**, 11-19.
[39] Fabijanowski, J. and Zarzycki, K. (1969) Spontane Vegetation als Grundlage für die Haldenaufforstung in Piaseczno bei Tarnobrzeg (Südospolen). *Beiheft zu den Zeitschriften des Schweizerischen Forstvereins*, **46**, 271-280.
[40] Bednarek, R. and Prusinkiewicz, Z. (1997) Soil geography (Geografia Gleb). PWN, Warsaw.
[41] Obmiński, Z. (1970) Zarys ekologii. in: Białobok, S., Ed., *Nasze Drzewa Leśne*, *Monografie Popularnonaukowe*, *Sosna Zwyczajna* (*Pinus silvestris L.*), Wydawnictwo Naukowe PWN.) Warszawa-Poznań, 203-231.
[42] Ellerbrock, R.H., Höhn, A. and Gereke, H.H. (1999) Characterization of soil organic matter from a sandy soil in relation to management practice using FT-IR spectroscopy. *Plant and Soil*, **213(1-2)**, 55-61
[43] Odum, H.T. and Odum, B. (2003) Concepts and methods of ecological engineering. *Ecological Engineering*, **20(5)**, 339-361
[44] Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., Van der Putten, W.H., Wall D.H. (2004) Ecological Linkages between aboveground and underground biota. *Science*, **304(5677)**, 1629-1633
[45] Shrestha, R.K. and Rattan L. (2006) Ecosystem carbon budgeting and soil carbon sequestration in reclaimed mine soil. *Environment International*, **32(6)**, 781-796
[46] Lieth, H. and Whittaker, R.F. (1975) Primary productivity of the biosphere. Springer Verlag, Berlin-Heidelberg-New York.
[47] Miller, A.T., Allen, H.L. and Maier, C.A. (2006) Quantifying the coarse-root biomass of intensively managed loblolly pine plantations. *Canadian Journal of Forest Research*, **36**, 12-22.
[48] Assmann, E. (1970) The principles of forest yield study: Studies in the organic production, structure, increment, and yield of forest stands. Pergamon Press, Oxford.
[49] Krebs, C.J. (2001) Ecology. The Experimental Analysis of Distribution and Abundance. Wydawnictwo Naukowe PWN, Warsaw.
[50] Orzeł, S., Socha, J., Forgiel, M. and Ochał, W. (2005) Biomass and annual production of mixed stands of the Niepołomice Forest (in Polish, English summary). *Act. Sci. Pol., Silv. Colendar. Rat. Ind. Lignar.*, **4(2)**, 63-79.
[51] Pietrzykowski, M. (2005) Characteristics of selected features of arborescent vegetation in reclaimed areas and in areas left for succession as exemplified by experimental plots in the Szczakowa sand mine excavation (in Polish, English summary). *Acta Agr. Silv. Ser. Silv.*, **63**, 1-26.
[52] Roberts, J.A., Daniels, W.L., Bell, J.C. and Burger, J.A. (1988) Early stages of mine soil genesis in a southwest Virginia spoil lithosequence. *Soil Science Society of America Journal*, **52**, 716-723
[53] Brożek, S. and Zwydak, M. (2003) Atlas of forest soils in Poland (in Polish). Atlas Gleb Leśnych Polski, CILP Press, Warsaw.
[54] Marrs, R. and Bradshaw A.D. (1993) Primary succession on man-made wastes: The importance of resource acquisition. In: Miles, J. and Walton, D.W.H., Eds., *Primary Succession of Land*, Blackwell Scientific Publications, Oxford, 221-247.
[55] Puchalski, T. and Prusinkiewicz, Z. (1975) Ecological background for forest site classification (in Polish). PWRiL Press, Warsaw.
[56] Harmsen, G.W. and Kolenbrander, G.J. (1965) Soil inorganic nitrogen. In: Bartholomew, W.V. and Clark, F.E., Eds., *Soil Nitrogen*, American Society of Agronomy, Madison, 43-92.
[57] Rumpel, C., Kögel-Knabner, I. and Hüttl, R.F. (1999) Organic matter composition and degree of humification on lignite-rich mine soils under a chronosequence of pine. *Plant and Soil*, **213(1-2)**, 161-168.
[58] Dobrzański, B. and Zawadzki, S. (1995) Pedology (Gelboznawstwo) (in Polish). PWRiL Press, Warsaw.
[59] Braun-Blanquet J. (1964) Pflanzensoziologie, Grundlage der Vegetationskunde (3 Aufl). Springer, New York.
[60] Ellenberg, H. (1979) Zeigerwerte der Gefäßpflanzen Mitteleuropas. *Scripta Geobotanica*, **9**, 7-122.
[61] Pietrzykowski, M. and Krzaklewski, W. (2006) Functionality assessment of soil quality index (ITGL) and vascular plant ecological indicator for site quality diagnosis in a sand mine excavation. *Acta Sci. Pol. Silv. Colendar. Rat. Ind. Lignar*, **5(1)**, 47-56

**Vol.2, No.6, 600-611 (2010)**
doi:10.4236/ns.2010.26075

**Natural Science**

# Evolution of technogenic landscapes by the example of apatite-nepheline ore concentration wastes

**Vladimir N. Pereverzev[1], Galina A. Evdokimova[2*], Irina V. Zenkova[2], Maria V. Korneykova[2], Vera V. Redkina[2]**

[1]Laboratory of Soil Science, Polar-Alpine Botanical Garden Institute, Kola Science Centre of Russian Academy of Sciences, Apatity, Russia
[2]Laboratory of Microorganisms Ecology, Institute of the North Industrial Ecology Problems, Kola Science Centre of Russian Academy of Sciences, Apatity, Russia; *Corresponding Author: galina@inep.ksc.ru

Received 10 February 2010; revised 30 April 2010; accepted 13 May 2010.

## ABSTRACT

**A primary soil-forming process can take place on the concentration waste of apatite-nepheline ores, whose biological recultivation was carried out more than 40 years ago. This process is characterized by the following features: forming of a thin litter with the content of organic carbon at the level of 8-12%, accumulation of humic substances in the sub-litter layer and the change of pH values. Microorganisms are biocatalysts of primary soil formation processes and one of the main factors that determine the specificity of this process. The prokaryotic complex of the newly formed soils, generated from nepheline sands, is considerably different from that of zonal soils on moraine sediments. The former ones are dominated by gram-positive bacteria, mainly actinobacteria, as well as by their filamentous forms (actinomycetes), whereas the latter ones are dominated by gram-negative bacteria. A common feature of invertebrate's complexes in nepheline sands is the low species diversity, small-size and quickly development of microfauna and mesofauna representatives and the dependence of succession of microarthropods pioneer groups on the succession of bacteria and fungi.**

**Keywords:** Soil-Forming Process; Nepheline Sands; Organic Matter; Soil Biota

## 1. INTRODUCTION

The mining and processing industry, which in the Murmansk region comprises several large enterprises, damages natural landscapes both as a result of open mining operations and due to the generation of overburden and mineral concentration waste dumps. As a result of activity of large concentration mills a large number of ore processing wastes are generated and stored in tailing dumps. The total area of territories covered by the tailing dumps in the Murmansk region at present makes about 5 thousand hectares. The operating and dormant tailing dumps represent a source of significant dusting of adjacent areas, since erosion processes are much developed on them as well as in natural and man-induced deserts. On the other hand tailing dumps can be considered as man-induced deposits that can be used in the future for recovery of valuable elements as soon as new techniques are introduced.

The investigation of primary soil formation processes, which take place on nepheline sands, was carried out from May to October 2005-2008 every years on tailing dumps with different duration of waste storage: 0, 10, 20, 30, 40 years. As a whole 180 sand samples have been taken for chemical and microbiological analyses and 300 samples for the zoological one.

## 2. NEPHELINE SANDS AS OBJECT OF BIOLOGICAL RECULTIVATION

Nepheline sands, as wastes of apatite-nepheline ores concentration, represent a soil-forming rock which is untypical for cold damp conditions. Nepheline sands are similar to the widespread in the region moraine, sea and fluvioglacial sands, on which dominating Al-Fe-humus podzols were formed, only by their granulometric composition. As well as in other soil-forming rocks, the mineral bulk of nepheline wastes comprise of fine sand fractions (0.25-0.05 mm – 29-31% in a layer of 0-15 sm) and coarse dust (0.05-0.01mm – 56-59%) with insignificant content of slime particles (< 0.001 mm –

2.4-3.0%).

A peculiarity of granulometric composition of nepheline sands is its heterogeneity with depths which is a consequence of influence of different conditions of deposit accumulation in the course of tailing dump filling.

By their bulk chemical composition nepheline sands significant differ from natural quaternary deposits (**Table 1**). The content of $SiO_2$ in them can be an indicator of high-, or low-grade soil-forming rocks. It is natural that the less the content of this element in rock, the higher the content of other elements including the biogenic ones, which determine the direction and intensity of biogeochemical processes in ecosystems. Sandy rocks with content of $SiO_2$ of 80-85% belong to high grade rocks.

In the Murmansk region the most widespread sea and continental rocks represented mainly by sands and sandy loams, differ from similar rocks of other northern regions by higher grade of chemical composition with the $SiO_2$ content of 65-67%. Eluvium of nepheline syenites on which the soils of tundra belt of the Khibiny were generated is characterized by even higher grade composition ($SiO_2$ < 60%). The composition of nepheline sands usually includes about 40% of $SiO_2$. They belong to very high grade rocks in their chemical composition. It contains considerable stocks of biogenic elements—phosphorus and potassium. Since apatite-nepheline ore does not contain any quartz, Si in them is represented by silicates, by basically nepheline ($NaAlSiO_4$). In the course of ore concentration the share of nepheline increases from 32 to 57%.

The content of P in nepheline sand is an order of magnitude higher than in moraine rocks since they contain apatite, not completely recovered in the process of ore concentration. Nepheline sands in terms of their supply with phosphorus also differ from cultivated soils both quantitatively, and qualitatively. They contain significantly more total phosphorus, but phosphorus is represented in them only by one composition—tricalcium phosphate (apatite).

**Table 1.** Total chemical composition of nepheline and moraine sand, percentage on ignited sample.

| Rock | $SiO_2$ | $Al_2O_3$ | $Fe_2O_3$ | CaO | MgO |
|---|---|---|---|---|---|
| Nepheline sands | 41.0 ± 0.6 | 21.0 ± 0.4 | 8.5 ± 0.3 | 6.5 ± 0.6 | 1.3 ± 0.1 |
| Moraine sands | 65.7 ± 1.3 | 13.8 ± 0.5 | 5.4 ± 0.3 | 4.1 ± 0.3 | 2.1 ± 0.2 |

| Rock | $TiO_2$ | $P_2O_5$ | MnO | $K_2O$ | $Na_2O$ |
|---|---|---|---|---|---|
| Nepheline sands | 2.6 ± 0.1 | 3.6 ± 0.6 | 0.18 ± 0.01 | 4.9 ± 0.1 | 10.7 ± 0.4 |
| Moraine sands | 0.9 ± 0.1 | 0.4 ± 0.1 | 0.12 ± 0.01 | 2.2 ± 0.3 | 3.7 ± 0.1 |

In soil, however, besides apatite, numerous other mineral compositions of phosphorus are present and, besides, some part of phosphates is a part of organic compositions [1,2]. Apatite is weakly soluble at the impact of soil solutions, therefore, despite the considerable content of phosphorus in nepheline sands, the cultivation of perennial grasses on them is impossible without introducing phosphoric mineral fertilizers.

The amount of potassium in tailings also exceeds that in moraine. On the average it makes 5%, which exceeds 2 times the content of this element in the cultivated podzolic soils [3].

Nepheline sands as a object of biological recultivation differ from zonal soils in the condition cation exchange capacity and the acid-base characteristics. If all soils generated on quaternary sediments are characterized by an acid medium, the nepheline sands have pH values in the alkaline range both in water, and salt suspensions (**Table 2**). In the course of long interaction of sands with the vegetative cover in the top part of mineral profile (to the depth of 20 sm) the reaction of medium in salt suspension changes into the acid range. While in deeper layers it remained alkaline. In water suspension the reaction of medium was alkaline at all depths. Other forms of acidity—the hydrolytic and the exchange ones—are also characterized by low indices not typical for zonal soils.

Thus, the reaction of nepheline sands fundamentally differs from that of zonal soils. At the same time the distribution of pH values both in sands and in soils, which follows the general law—with depth the reaction of medium shifts towards neutral or alkaline values. It is in that the influence of eluvial processes tells on the mineral profile.

Initial nepheline sands are devoid of organic substance of biological origin. Presence of organic carbon in them is due to the remains of flotation reagents—a mixture of resin and fatty acids used in the technological process during that period. They are rather stable in time, which is confirmed by the presence of organic carbon in nepheline sands of 20-30-years "age". The content of organic carbon of technogenic origin in sands makes 0.2-0.3%.

**Table 2.** Average pH values in nepheline sands and podzols.

| Depth, sm | pH ($H_2O$) | | pH (KCl) | |
|---|---|---|---|---|
| | Soil on sands | Podzol | Soil on sands | Podzol |
| 0-1 | 7.1 | 4.3 | 5.8 | 3.5 |
| 1-5 | 7.5 | 4.7 | 6.2 | 3.8 |
| 5-10 | 7.6 | 5.1 | 6.4 | 4.6 |
| 20-60 | 8.2 | 5.5 | 7.1 | 5.0 |

Thus, nepheline, being the basis of mineral bulk of sands, is an unstable mineral which is exposed to intensive weathering under the impact of acid solutions, including humus acids. The extremely high mobility of biogenous elements which are a part of nepheline sands is connected with this [4]. The soil-forming rocks on which zonal soils of the Murmansk region were formed do not have similar properties in such pronounced form. On the other hand, the absence of organic substance and fixed nitrogen in nepheline sands, as well as the inaccessibility of large reserves of phosphorus to plants requires solution of problems of optimization of the nutritive status of sands when growing on them plants for the nature protection and eventual economic purposes. Nepheline sands are suitable for cultivation of plants not only with a view of their fixation from wind erosion, but also for creation of productive agricultural lands, which proved true when carrying out field pilot works after growing meadow grasses on a fixed tailing dump [5].

## 3. NEWLY FORMED SOILS ON NEPHELINE SANDS

The studies of the structure and properties of soils formed on vegetation fixed nepheline sands ware carried out on dormant tailing dump. Recultivation took place from 1964 to 1984 by sowing perenial grasses. Over a part of the tailing dump, recultivated in 1964-1968, a vegetative cover of various structures was generated depending on edaphic conditions, first of all, apparently, on the character of substrate humidification. Along with cereal grasses, some shrubs and subshrubs, motley grasses, in particular, red clover, participate in the formation of the ground cover, as well as mosses and lichens, typical for zonal phytocenoses. The wood canopy is represented by rare specimens of pine and spruce, planted in 1978 in the course of biological recultivation [6], and heavy birch thicket of 20-30-years age, which penetrated in the phytocenosis by natural seeding after the tailing dump surface fixation using perennial grasses. Some rare specimens of alder and juniper are found. Glades, devoid of tree and shrubs and herbaceous layer have a moss-lichen cover continuum.

The soils inherited their total chemical composition from soil-forming rocks. The newly cultivated soils which age is estimated at several decades, can not essentially differ by their chemical composition from rocks on which they are formed. The rich chemical composition peculiar to rock is also characteristic for newly formed soils. The average for all sections content of $SiO_2$ in soils on nepheline sands is 41% on ignited soil while in the arable layer of cultivated podzols it exceeds to 65%.

The total content of $Fe_2O_3$ and $Al_2O_3$ reaches almost 30%, alkaline-earth bases—7.8%, alkaline metals—15.6%, of them the share of $Na_2O$ is 10.7%. As said above, nepheline sands contain a lot of phosphorus as a part of apatite, which remained in tailings after the ore concentration. The reserves of phosphorus in soil, naturally, have been preserved. In nepheline sands and in soils generated on them, a high enough content (0.34% on average) of fluorine had been registered. In zonal podzols the content of this element usually made no more than 0.01-0.2% [7]. Fluorine is a part of apatite composition (a variety—fluorine-apatite) which is the reason of enrichment of nepheline sands with it. Thus the initial soil formation on nepheline sands proceeds in conditions of very rich chemical composition of soil-forming rocks.

### 3.1. Organic Matter of Newly Formed Soils

Formation humus substances, specific organic compounds that are peculiar to soils, are the initial stage of soil profile formation. As a result of transformation by microorganisms of the vegetative litter on the surface of nepheline sands an organic horizon was generated. In connection with low power (0.5-1.5 sm) it was enriched by mineral particles. Therefore the content of organic carbon in it is relatively small—8-11%.

The average data of the content of organic carbon based on all investigated ecotopes are presented in **Table 3**. The greatest spatial variability of organic carbon value in organic horizon is noted (the variation coefficient makes 46%). In the top layers of mineral mass (in cespitose horizon and at the depth of 5-10 sm) the variability is insignificant (< 10%) and mean (10-20%) [8]. By this reason the difference in the content of organic carbon in these horizons is reliable. The difference of mean values was 0.2%, the error of difference was 0.031% and the Student criterion was 6.5 ($t_{05}$ = 2.31). Hence, we can assert with sufficient reliability that under the influence of biota, which transforms plant residues, in the formed cespitose horizon an accumulation of organic substance took place in amounts exceeding its content in the initial nepheline sands.

**Table 3.** The average content of organic carbon.

| Horizon (Depth, sm) | Organic C, % | Coefficient of variation, % |
|---|---|---|
| AO (mulch) | 7.66 ± 1.445 | 46 |
| AY (cespitose) | 0.54 ± 0.025 | 9 |
| 5-10 | 0.34 ± 0.019 | 13 |
| 10-20 | 0.31 ± 0.029 | 23 |
| 20-30 | 0.28 ± 0.025 | 22 |
| 30-40 | 0.28 ± 0.027 | 23 |

The content of water-soluble compounds of carbon, which is a part of organic substance of soils as a whole, is closely connected with the content of total carbon. To a greater extent, it concerns the organic horizon of soils, where rather a close correlation dependence of amount of water-soluble and total carbon ($r = 0.728 \pm 0.278$ $t = 2.81$; $t_{05} = 2.18$) is found. In mineral layers of soils this connection is lesser significant ($r = 0.384 \pm 0.161$; $t = 2.38$; $t_{05} = 2.00$). Probably, it is connected with the fact that in the mineral mass of sands the organic substance is represented both by compounds of biological nature and by nonspecific compounds—the remains of flotoreagents.

In profile distribution of indices of water-soluble carbon the same regularities, as in distribution of the total carbon are manifested (**Table 4**). It is natural, that the most of all water-soluble organic substance is found in organic horizon. In mineral horizons its amount decreases rapidly, but in the cespitose layer (up to 5 sm) the content of water-soluble carbon is higher, than in the underlaying layers.

**Figure 1** presents the average data on all ecotopes, illustrating the regularities of profile distribution of the total and water-soluble carbon.

Their absolute content follows the general regularity: rapidly decrease in direction from organic horizon to the cespitose one and the further gradual decrease with depth. Concerning the values of the degree of mobility of organic substance (of the content of water-soluble carbon in percents of the total) the pattern is reverse. In direction from organic horizon to the cespitose one this value increases and continues to increase in the deeper layers. Hence, in the process of decrease in the content of organic substance with depth its relative mobility increases. The registered regularities are well described by exponential curves with rather a high approximation reliability ($R^2 = 0.88$-$0.90$).

By the fractional structure of humus acids soils on nepheline sands differ from zonal soils: in the composition of humic and fulvic acids the fractions connected with calcium is played an appreciable role. It is caused by abundance of bases in nepheline sands, including the ones in mobile condition.

## 3.2. Acid-base Properties of Newly Formed Soils

Acidity is an important indicator of soils fertility, formed as a result of interaction of plants with soil-forming rock, which in our case is represented by nepheline sands. Initial sands have alkaline reaction both in water (pH 8.0-8.3) and in salt (pH 7.3-7.8) suspensions.

In the prosess of primary soil formation on the surface of nepheline sands covered with plants, a thin organic horizon there was generated, in which processes of transformation of plant litter take place, resulting in the formation of organic acids which interact with mineral mass of sands. The result of such interaction is accumulation of newly formed soil of organic mineral compounds being of acid nature in the top part of a mineral profile. The results of pH determination in water suspension show, that an organic horizon has, as a rule, neutral reaction, and in a salt suspension the reaction of this horizon becomes subacidic (**Table 5**).

**Table 4.** The average content of water-soluble carbon.

| Horizon (Depth, sm) | Water-soluble C, mg/100 g | Coefficient of variation, % |
|---|---|---|
| AO | 99 ± 10.8 | 26.9 |
| AY (cespitose) | 18 ± 0.9 | 9.4 |
| 5-10 | 11 ± 2.2 | 49.8 |
| 10-20 | 9 ± 2.1 | 57.9 |
| 20-30 | 11 ± 1.9 | 43.5 |
| 30-40 | 11 ± 2.7 | 59.4 |

**Table 5.** Average pH values in nepheline sand and podzols.

| Depth, sm | pH ($H_2O$) | | pH (KCl) | |
|---|---|---|---|---|
| | Soil on sands | Podzol | Soil on sands | Podzol |
| 0-1 | 7.1 | 4.3 | 5.8 | 3.5 |
| 1-5 | 7.5 | 4.7 | 6.2 | 3.8 |
| 5-10 | 7.6 | 5.1 | 6.4 | 4.6 |
| 20-60 | 8.2 | 5.5 | 7.1 | 5.0 |

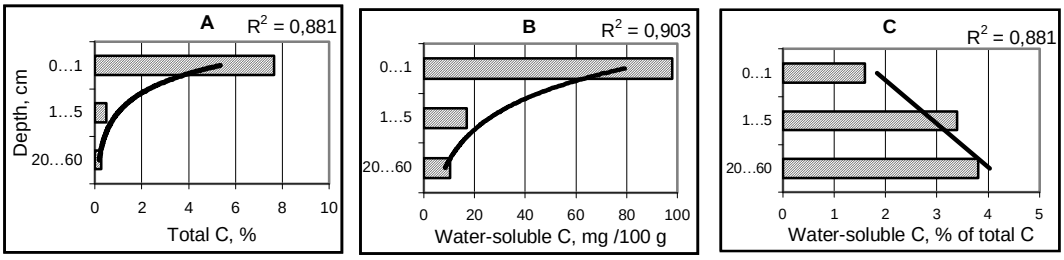



**Figure 1.** Distribution of total carbon over soil profile (A); % of water-soluble carbon, mg/100 g (B); water-soluble carbon, % of total carbon (C). Average data.

In a profile of nepheline sands the medium response changes with depth from neutral to alkaline in water suspension and from subacidic to neutral in salt suspension. In zonal Al-Fe-humus podzols on sandy sediments along the entire profile acid response both in salt and in water suspensions is observed. The distribution of pH values in sand and soils follows the general regularity—with the depth the medium response shifts towards neutral or alkaline values. It is through this that the influence of eluvial processes affects the mineral profile.

Thus, the newly formed soils, generated on nepheline sands, differ considerably by the content and distribution of organic substance from zonal soils—Al-Fe-humus podzols. The presence of organic horizon, which is less mineralized in podzols, even in case of small thickness is a common feature. Differences in the content in this horizon of general and water-soluble carbon are connected with it: there is 4 times more total carbon in zonal soils than in soils on nepheline sands, and 10 times more of water-soluble carbon. By the content of the total carbon the cespitose horizon of newly formed soils is similar to podzolic horizon of podzols. These horizons, along with litter, in the both soils are horizons with greatest concentration of root systems of plants. No illuvial accumulation of organic substance, which is characteristic for podzols, is taking place in soils on nepheline sands, possibly, owing to their youth.

A long (30-40-years) influence of vegetation on the nepheline sands has neither caused any appreciable change in the total content of macroelements nor has it led to their differentiation within a mineral profile.

## 4. SOIL BIOTA

Soil biota is a biocatalyst of processes of primary soil formation. It participates directly in the mineralization and humification of plant residues, affecting them by its hydrolytic and oxidation-reduction exoenzymes, mineral and organic acids and other metabolites. The investigation of biological characteristics was carried out in sands of different duration of storage in tailing dumps apatite-nepheline processing plants.

### 4.1. The Number and Biomass of Bacteria

The tailings of apatite-nepheline ore concentration leaving the plant are not sterile and contain up to $n \times 10^7$ of bacteria in 1 ml of discharge [9]. Their amount decreases up to $n \times 10^6$ of cells in 1 g of sand at their ingress into the tailing dump, which is most often connected with the low humidity of the habitat and a decrease in concentration of nutritious elements (**Table 6, Figure 2**). The intensity of microbiological processes increases in the process of planting higher plants on sands.

**Table 6.** Number of bacteria of various trophic groups ($\times 10^3$ cell/g) in nepheline sands.

| Sample, plant group | "Age" of sand, years | Saprotrophic bacteria | Using mineral N | Oligotrophic bacteria |
|---|---|---|---|---|
| Pure sand | 0 | 130 – 3310<br>1394 ± 213 | 10 – 6200<br>1386 ± 319 | 0 – 2630<br>720 ± 162 |
| | 10 | 480 – 9980<br>1875 ± 464 | 170 – 7320<br>2000 ± 382 | 120 – 5910<br>1440 ± 286 |
| | 20 | 120 – 1580<br>741 ± 97 | 20 – 1660<br>594 ± 109 | 10 – 2130<br>680 ± 116 |
| | 30 | 0.7 – 169<br>40 ± 8 | 0 – 360<br>55 ± 20 | 0 – 280<br>78 ± 19 |
| Grass | 10 | 970 – 16100<br>4495 ± 760 | 310 – 10860<br>3776 ± 664 | 10 – 20500<br>5551 ± 1279 |
| | 20 | 680 – 9060<br>3163 ± 452 | 100 – 21300<br>3334 ± 991 | 300 – 19800<br>3126 ± 1032 |
| | 30 | 790 – 17700<br>5752 ± 953 | 0 – 20200<br>5664 ± 1299 | 300 – 17900<br>4966 ± 1070 |
| Moss | 40 | 1860 – 15200<br>5491 ± 812 | 300 – 14400<br>5439 ± 1108 | 300 – 16600<br>6417 ± 1214 |
| Lichenous | 40 | 2300 – 17600<br>6176 ± 692 | 500 – 18700<br>5394 ± 1067 | 720 – 15200<br>3913 ± 741 |
| Motley-grass | 40 | 2480 – 25000<br>10382 ± 1388 | 400 – 67700<br>17714 ± 3828 | 900 – 82700<br>18920 ± 5258 |
| Shrub | 40 | 1280 – 19200<br>5328 ± 903 | 700 – 11100<br>3782 ± 611 | 1100 – 12300<br>4734 ± 612 |

Note. Above line – lim (min-max), under line - M ± m; n = 25-28.

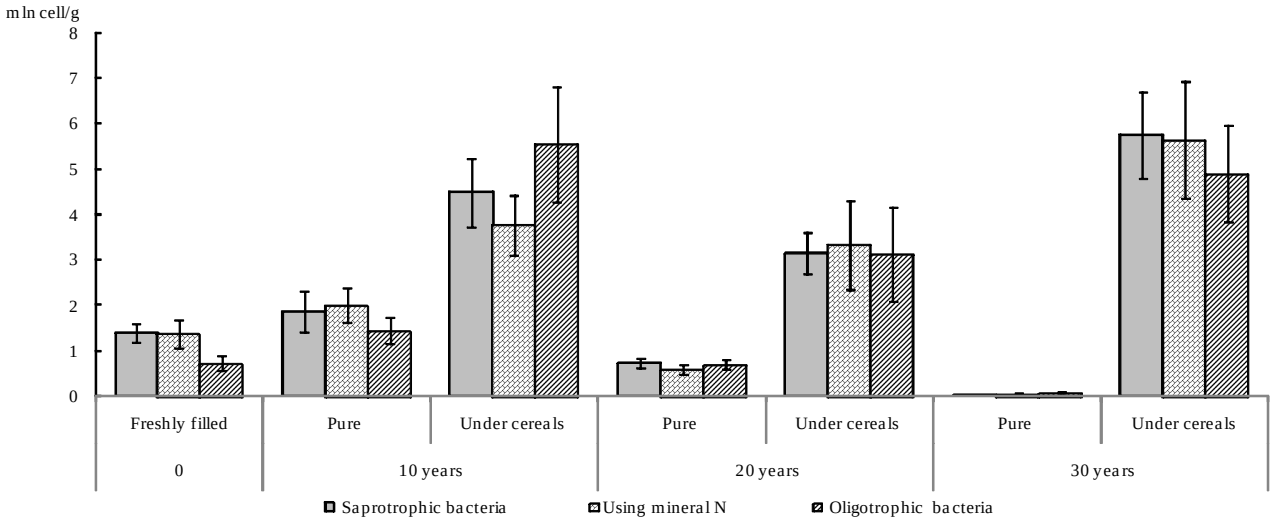



**Figure 2.** The number of bacteria in non-recultivated nepheline sands of different storage duration and in sand under cereals.

The highest number of all the trophic bacterial groups was reached under grass parcels.

So, in a microbocenos developed in the thin organic horizon, generated at a nepheline tailing dump revegetated 40 years ago, the leading position in which was occupied by oligotrophic bacteria and bacteria transforming complex organic nonnitrogenous substances, in particular, representatives of amylolytic community. The number of all trophic groups of bacteria was the greatest under of grass parcel.

More complete data about the number of microorganisms in soils can be obtained using microscopic counts methods, in particular, the method of fluorescent microscopy. Based on the data of the total content of bacteria in the substrate, it is possible to calculate their biomass.

The total number of bacteria in pure sand using the method of fluorescent microscopy, which considers both viable and unviable cells, varied within 0.34-0.60 billion cell/g, and in the recultivated one under various plant gropes—from 5.8 to 7.2 billion cell/g (**Table 7**). The "age" of pure, unrecultivated sand practically did not exert any influence on the total number of bacteria and their biomass. In sands under grass plants the bacterial biomass increased on the average 4 times compared to the sand which was not covered with plants. At that, no reliable changes in the amount of bacteria biomass occur as the age of the sand increases. The greatest number and biomass of bacteria was under motley grasses, dominating in which was clover, capable to symbiotic fixation of nitrogen.

Thus, in the recultivated sands the bacterial biomass has increased on the average 14 times in comparison with the sands not covered with vegetation, and changed under various plants groups within 0.11-0.29 mg/g.

## 4.2. Peculiarities of the Prokaryotic Complex of Newly Formed Soils

The prokaryotic complex of newly formed soils on nepheline sands essentially differs from the prokaryotic complex of zonal soils on moraine. In the prokaryotic complex of the studied substrate gram-positive bacteria dominate, whereas in zonal soils gram-negative bacteria prevail which testifies to differences in the species composition of the bacterial community. In non-recultivated sands the share of gram-negative bacteria changed from 4 to 10% of the total number of organotrophic bacteria. In the recultivated sands their share

**Table 7.** The total number of bacteria ($\times 10^9$ cells/g) and their biomass ($\times 10^{-5}$ g/g) in nepheline sands.

| Sample, plant group | "Age" of sand, years | Number | Biomass |
|---|---|---|---|
| Pure sand | 0 | 0.60 ± 0.04 | 2.4 ± 0.2 |
| | 10 | 0.53 ± 0.06 | 2.1 ± 0.3 |
| | 20 | 0.34 ± 0.11 | 1.4 ±0. 4 |
| | 30 | 0.59 ± 0.10 | 2.4 ± 0.4 |
| Grass | 10 | 1.47 ± 0.37 | 5.9 ± 0.9 |
| | 20 | 2.23 ± 1.04 | 8.9 ± 0.7 |
| | 30 | 2.02 ± 0.19 | 8.1 ± 0.6 |
| Moss | 40 | 7.22 ± 0.37 | 28.9 ± 1.5 |
| Lichenous | 40 | 5.79 ± 0.64 | 23.2 ± 2.5 |
| Motley-grass | 40 | 6.20 ± 0.27 | 24.8 ± 1.1 |
| Shrub | 40 | 2.62 ± 1.64 | 10.5 ± 6.6 |

increased to 30-50% (**Table 8**).

A distinctive feature of a microbic component of the newly formed soils, generated on nepheline sands, from acid soils of the region on moraine sediments, was the high number of actinomycetes of genus *Streptomyces*, class Actinobacteria. In forest podzols their amount does not exceed 3.5% of the total number of saprotrophic bacterial block, while in the recultivated tailing dump streptomycetes reach up 25% of the total number of cultivated bacteria. The representatives of genus *Nocardia* of the same class s were found much less often.

Actinomycetes are neutrophils, while water suspensions of nepheline sands possess neutral or alkaline reaction, at the same time for Al-Fe-humus podzols dominating in the Kola Peninsula, acid reaction of the medium is characteristic. Actinomycetes produce extracellular hydrolases, capable of decomposing complex organic compounds: cellulose, xylogen, chitin, humus substances.

As a whole the prokaryotic complex of newly formed soils on nepheline-bearing industrial wastes is presented in **Table 9**.

Its composition includes mainly actinobacteria of genera *Arthrobacter*, *Rhodococcus*, *Micrococcus* and *Streptomyces*, adapted for life in oligotrophic media as a result of economical consumption of both exogenous and endogenous substrates. The composition of prokaryotic complex of pure non-recultivated sand and the sand prior to 30-years old, which has overgrown with grasses and mosses, includes mainly actinobacteria of genera *Arthrobacter* and *Micrococcus*, often forming associative colonies on nutrient media (therefore, their share in the total complex of saprotrophic bacteria could exceed 100%).

**Table 8.** The share of gram-negative bacteria in nepheline sands (% of the total number of saprotrophic bacteria).

| Sample, plant group | "Age" of sand, years | Share of Gr $^-$ bacteria |
|---|---|---|
| Pure sand | 0 | 4 |
| | 10 | 5 |
| | 20 | 10 |
| | 30 | 4 |
| Grass | 10 | 21 |
| | 20 | 12 |
| | 30 | 11 |
| Moss | 40 | 41 |
| Lichen | 40 | 31 |
| Motley-grass | 40 | 52 |
| Shrub | 40 | 39 |

**Table 9.** Prokariotic complex of nepheline sands of different period exposition (% from organotrophic bacterial block).

| Sample, plant group | Total number, $\times 10^6$ cells/g | *Arthtobacter* | *Chryseobacterium* | *Rhodococcus* | *Micrococcus* | *Streptomyces* | *Bacillus* |
|---|---|---|---|---|---|---|---|
| Pure sand | 1.0 | 59.6 | 0.2 | 12.0 | 46.8 | 3.7 | 0.7 |
| Grass (10–30 years) | 4.5 | 63.8 | 0.9 | 0 | 32.8 | 8.4 | 1.1 |
| Moss (10–30 years) | 2.2 | 69.9 | 0 | 0 | 31.7 | 4.8 | 0.9 |
| Moss (40 years) | 5.5 | 20.1 | 0.8 | 0 | 0 | 25.8 | 7.4 |
| Lichen (40 years) | 6.2 | 18.0 | 1.0 | 0 | 0 | 25.8 | 2.0 |
| Motley-grass (40 years) | 10.4 | 10.0 | 0.8 | 0 | 0 | 13.5 | 3.8 |
| Shrubs (40 years) | 5.3 | 10.8 | 2.4 | 0 | 3.4 | 17.6 | 2.3 |

In nepheline sands 5 strains of dominating species of bacteria have been secured with more than 60% of spatial frequency of occurrence. Their identification has been carried out using the method of comparative analysis of nucleotide sequences of genes, coding 16S rRNA, and their phylogenetic position ("Bioengineering" centre, Moscow) has been determined. Four strains of the identified bacteria have been referred to *Actinobacteria* class. These are strains of species*: Arthrobacter boritoleran*, *A. ramosus*, *Rhodococcus fascian*, *Micrococcus luteus.*

Actinobacteria are typical dwellers of soils, water, air and are characterized by non-specificity to nutrient sources and can develop in media with the small content of nutrients, thanks to economical consumption of exogenous substrates, *i.e.*, they belong to the trophic group of oligotrophic bacteria. Some of them, in particular, actinomycetes are capable of producing extracellular hydrolases and of decomposing complex polymeric compounds. Besides, actinobacteria can develop at very low humidity of substrate and have high adaptive capacities to adverse conditions of the environment in particular they form carotinoids, protecting cells from UV rays.

During long-time exposition in the process of formation of newly formed soils on nepheline-bearing industrial wastes the structure of prokaryotic complex of microbial communities was changes. The share of gram-negative bacteria increases in them from 4-10% to 30-50%, while the share of actinobacteria, belonging to gram-positive bacteria, including streptomycetes decreases. The domination of gram-negative bacteria in prokaryotic complex and exclusively small content of actinomycetes is characteristic of acid soils of the region on moraine sediments [10]. Based on this we can assume, that evolution of microbic community of nepheline sands in process of their recultivation and revegetation follows

the way of rapprochement with microbial communities of zonal soils.

### 4.3. The Number and Biomass of Fungi

The number of microscopic fungi—the basic decomposers of organic substance in the recently filled sands from apatite-nepheline manufacture was very small and did not exceed tens of CFU per 1 gram (**Figure 3**). In the process of increase of the storage period of sands and their revegetation the number of fungi increased to hundreds of CFU per gram of substrate.

More complete data about the abundance of fungi in soils can be obtained using microscopic counts methods, in particular, the method of fluorescent microscopy. The length of fungal mycelium in nepheline sands, which had been object to recultivation 40 years ago, under lichen and motley grass group reached 1000 m/g, and its biomass made 1.3 mg in 1 g of substrate (**Table 10**). These values are quite comparable to those in soils of taiga forests of the Kola Peninsula [10]. At that, the fungi biomass in the 40-year sands exceeded the bacterial

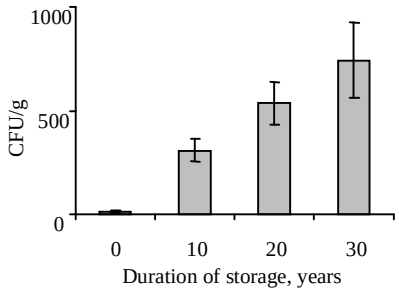



**Figure 3.** Number of microscopical fungi in nepheline sands in the time gradient.

**Table 10.** The length of fungal mycelium (m/g) and the biomass of fungi and bacteria ($\times 10^{-5}$ g/g) in nepheline sands.

| Sample, plant group | "Age" of sand, years | Length of fungal mycelium | Fungal biomass | Total biomass of fungi and bacteria |
|---|---|---|---|---|
| Pure sand | 0 | 12 ± 5 | 1.3 ± 0.6 | 3.7 ± 0.4 |
| | 10 | 32 ± 2 | 3.5 ± 0.2 | 5.6 ± 0.2 |
| | 20 | 43 ± 13 | 4.7 ± 1.5 | 6.1 ± 0.9 |
| | 30 | 26 ± 5 | 2.9 ± 0.5 | 5.3 ± 0.4 |
| Grass | 10 | 59 ± 8 | 6.4 ± 0.9 | 12.3 ± 0.7 |
| | 20 | 62 ± 8 | 6.9 ± 0.8 | 15.8 ± 0.7 |
| | 30 | 112 ± 35 | 12.4 ± 3.9 | 20.5 ± 2.2 |
| Moss | 40 | 710 ± 145 | 78.1 ± 16.0 | 107.0 ± 8.7 |
| Lichen | 40 | 1156 ± 192 | 127.2 ± 21.1 | 150.4 ± 11.8 |
| Motley-grass | 40 | 1064 ± 74 | 117.0 ± 8.1 | 141.8 ± 4.6 |
| Shrub | 40 | 434 ± 159 | 47.8 ± 1.7 | 58.3 ± 4.1 |

one 3-5 times which is characteristic of organic horizons of the zone of spruce forests of the Kola North. In the recently filled sands the indices of mycelium length and its biomass were much lower and did not exceed 12 m and 0.013 mg/g respectively.

It should be noted, that in sands without vegetation the contribution of bacteria and fungi to the total microbic biomass is equivalent, and in the sands subjected to phytomelioration, the fungi biomass exceeds 10 times that of bacteria.

### 4.4. Fungi Species Diversity

At present we have identified in the sands of the tailing dump, reclaimed over 40 years ago, —26 species related to 10 genera, 7 orders, 4 classes and 2 divisions; in recently filled sands—only 12 species related to 8 genera, 5 orders, 4 classes and 2 divisions. Most widely represented in the complex of micromycetes of the reclaimed nepheline tailing dump have been fungi of *Penicillium* genus. They made over 50% of all species diversity of the identified fungi. In recently filled sands the given genus was represented by 4 species, in the recultivation one – by 15 species.

In nepheline sands recultvated over 40 years ago, the group of often found fungi included the species: *Mortierella longicollis, Phoma eupyrena, Penicillium daleae.* Fungi *Acremonium rutilum, Fusarium solani, Mucor hiemalis, M. plumbeus, Penicillium variabile* have been identified only in recently filled nepheline sands. These species of fungi have been also found in apatite-nepheline underground mining workings [11] and in products of technological conversion at apatite-nepheline concentrating mills [9], whence they could go to the tailing dump.

No dominating species were found in recently filled sands, which are also confirmed by the decrease of the value of Simpson domination index and, respectively, the increase of the value of Pielou evenness index (**Table 11**).

In recultivated sands Simpson index was equal to 0.26, Pielou—0.53; in recently filled sands to 0.15 and 0.96

**Table 11.** Some indices of species structure of nepheline sand fungi community.

| "Age" of sand | Shannon total diversity | Simpson domination | Pielou evenness |
|---|---|---|---|
| 0 | 1.99 | 0.15 | 0.96 |
| 10 | 1.31 | 0.4 | 0.46 |
| 20 | 1.63 | 0.37 | 0.53 |
| 30 | 1.57 | 0.29 | 0.55 |
| 40 | 1.7 | 0.26 | 0.53 |

respectively. *Penicillium thomii* belonged to the frequently found species in recently filled sands, other species by the values of spatial and temporal frequency of species occurrence, to rare and casual ones have been referred. The registered low species diversity of micromycetes in recently filled sands and the absence of species—dominants in the structure of their complexes are characteristic for young ecosystems in unstable condition and they are fewer, where physical and chemical factors of the medium are extreme.

The degree of similarity of the species composition of complexes of microscopic fungi of recently filled and reclaimed nepheline sands, expressed by Sørensen index, has made only 25%. Such low degree of similarity is explained by the extremely small value of the number and poor fungi species diversity in the concentration wastes leaving the mill. For recultivated nepheline sands the value of Sørensen index is above 40%.

The similarity of species composition of complexes of micromycetes of recultivated nepheline sands and typical podzolic soils of the Kola Peninsula increases—the Sørensen index reached 45%.

Thus, the recultivation of nepheline sands, carried out 40 years ago, provided preconditions for formation of complexes of bacteria and micromycetes, typical for regional podzolic soils. In the formation of pioneer complexes of micromycetes of nepheline sands the process of restoration of vegetative cover determining their number and diversity is the most essential, while differences in mineralogical and chemical composition of sands, which are parent rock are not so significant.

## 4.5. Zoocenoses

Along with phyto- and microbocenoses, zoocenoses are an obligate component of a soil ecosystem. In soils, disturbed as a result of agrotechnical activity or industrial impact, microarthropods—mites and collembolans appear to be the main and often the only groups of fauna. Small sizes, high number, diversity of species and life forms, wide ranges of food resources determine the universal spread of microarthropods. As a rule, they are the first of animals which colonize technogenic substrates after microorganisms.

### 4.5.1. Diversity of Invertebrate Animals

Zoological analysis of nepheline sands has shown that invertebrates-colonizers of this man-induced substrate are collembolans (Insecta, Collembola). Colonization of sands by these microbophagous insects is explained by the presence of living bacterial cells in the mineral concentration waste of apatite-nepheline ore arriving to storage sites from mills. Accumulation of humus substances and succession of bacteria and fungi components of sands microbiota have determined the change of microarthropod´s dominating groups. In newly formed soils at the recultivated storage site the saprotrophic oribatid mites (Acari, Oribatei) as indicator of humification processes of organic substance were absolute dominants of fauna.

The storage site, recultivated over 40 years ago, was characterized by the greatest taxonomic and trophic diversity of invertebrates. The newly formed organic horizon was occupied by litter-dwelling animals with many-years generations and ecological strategy of *K*-type: detritophagous earthworms of the species *Lumbricus rubellus* (Lumbricidae), carnivorous millipedes of the species *Monotarsobius curtipes* (Lithobiidae) and hydrophilic larvae of leather-winged beetles (Cantharidae). In connection with plants diversity the complex of phytophagous insects has extended (**Table 12**).

**Table 12**. Diversity of trophic groups of invertebrates in nepheline sands with different period of exposition.

| Period of exposition, years | | | |
|---|---|---|---|
| 10 | 20 | 30 | 40 |
| *Microbophagous* | | | |
| Collembola | Collembola | Collembola | Collembola |
| Oribatei | Oribatei | Oribatei | Oribatei |
| | | Nematoda | Nematoda |
| | | Tartigrada | Tartigrada |
| *Saprophagous* | | | |
| Diptera | Diptera | Diptera | Diptera |
| Byrridae | Byrridae | | |
| | Enchytraeidae | | Enchytraeidae |
| | | | Lumbricidae |
| *Phytophagous* | | | |
| | Cicadellidae | Cicadellidae | |
| | Elateridae | Elateridae | Elateridae |
| | Thysanoptera | | Thysanoptera |
| | | Aphididae | Aphididae |
| | | | Lepidoptera |
| | | | Hemiptera |
| *Carnivorous* | | | |
| Mesostigmata | Mesostigmata | Mesostigmata | Mesostigmata |
| Aranea | Aranea | Aranea | Aranea |
| Staphylinidae | Staphylinidae | Staphylinidae | Staphylinidae |
| | Carabidae | Carabidae | Carabidae |
| | Cantharidae | | Cantharidae |
| | | Formicidae | Formicidae |
| | | | Lithobiidae |

Representatives of the majority of taxa, typical for taiga soils (nematodes, pot worms, spiders, road beetles, carabid beetles, leather-winged beetles and dipterans larvae) preferred plant associations of the forest type—mosses, lichens, shrubs and lived under them throughout the period of vegetation. Phytophagous insects (aphid, lemipterans, thunder flies, caterpillars of lepidopterans) were various under lichens. For certain plant associations the obligate groups of invertebrates there have been identified, which were found under them during all vegetative season. Under mosses, lichens, shrubs and clover these were nematodes and pot worms, under red whortleberries and crowberries—larvae of dipterans and road beetles.

### 4.5.2. Number of Invertebrate Animals

In the sands of operating tailing dump the total number of invertebrates varied from 0.5 to 17 thous. spm/m$^2$ (**Figure 4**). The mean by 6 months number of fauna without taking into account of microarthropods has not exceeded 30 spm/m$^2$ in pure sand of different period of exposition and 200-400 spm/m$^2$ under grasses and mosses. At the dormant tailing dump in newly formed soils the number of microarthropods has increased up to 47 thous. spm/m$^2$ and other groups of fauna up to 740 spm/m$^2$.

In the "young" sands, both self-growing with mosses and fixed by sowing grasses 10-30 years ago, the pioneer complexes of invertebrates are connected with the moss cover, rhizosphere and the overground biomass of grasses. In sands under mosses the tundra variant of invertebrates community was formed and under motley-grass—the meadow one. The greatest number and diversity of species and life forms of invertebrates is confined to grasses, which testifies to an important medium-forming activity of these plants.

The character of the change of number of microbophagous—collembolans and oribatid mites in sands of different age under all plant associations correspondents with dynamic of the number of microorganisms – bacteria and mycelial fungi (**Figures 5, 6**).

It follows from figures, that the succession of pioneer groups of microarthropods—collembolans and mites in the studied technogenic substrate is determined by the succession of microorganisms. A similar conclusion has been made at the analysis of zoo-microbial interactions during experimental field on biotransformation of plant residues in soils polluted by aluminium and copper-nickel smelters emissions in the Murmansk region [7, 12-14].

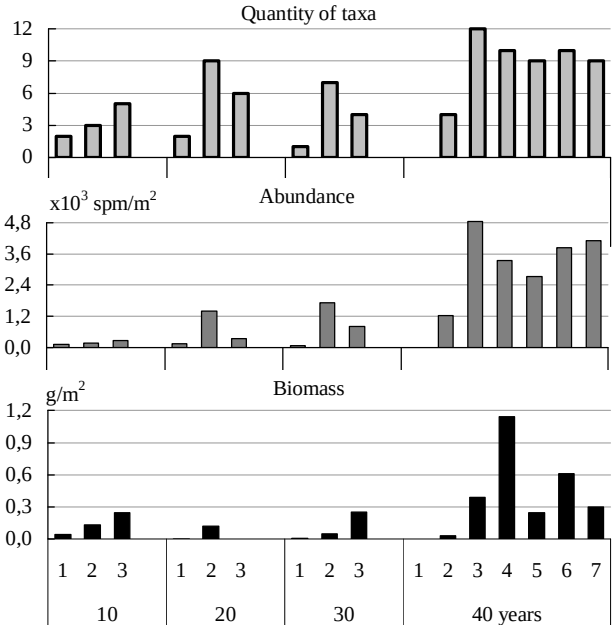



**Figure 4.** Quantity of taxa, number and biomass of invertebrates without microarthropods in the nepheline sands (mean on 6 season of sampling). 1—pure sand, 2—grasses, 3—mosses, 4—lichens, 5—crowberry, 6—red whortleberries, 7—clover.

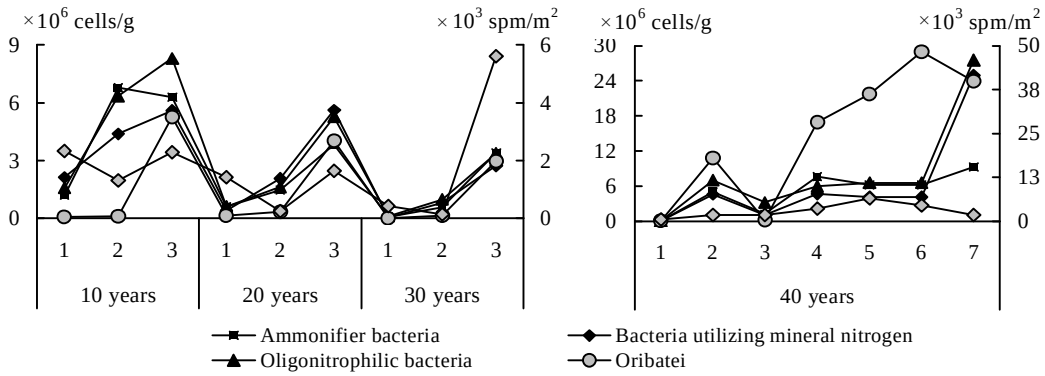



**Figure 5.** The number of bacteria (10$^6$ cells/g) and microarthropods (10$^3$ spm/m$^2$) in the sand with different period of exposition under the main types of vegetation. Legend 1-7 as well as on **Figure 4**.

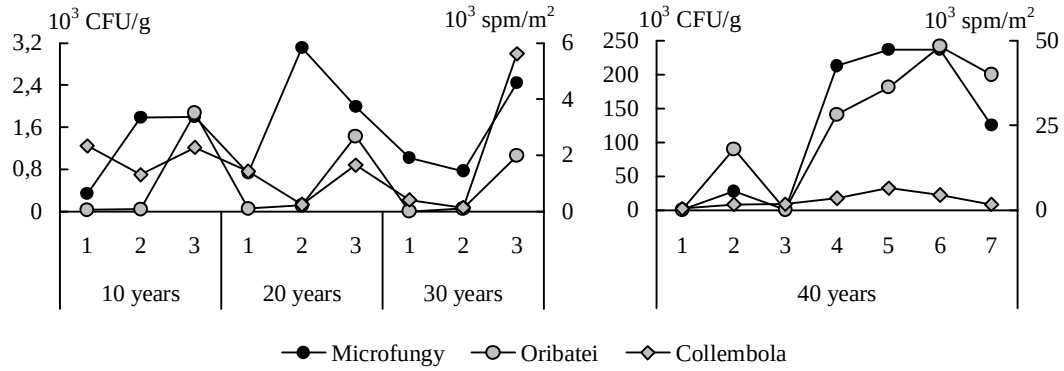



**Figure 6.** The number of microfungi ($10^3$ CFU/g) and microarthropods ($10^3$ spm/m$^2$) in the sands with different period of exposition. Legend 1-7 as well as on **Figure 4**.

Phytomelioration of nepheline sands has positively affected the growth of the number of fauna compared with the sands without vegetation or self-growing with mosses. As a whole, biotic factors have determinant influence on the colonization of this technogenic substrate by various groups of invertebrate animals in comparison with abiotic factors. The bacterial biomass in the newly formed soils on nepheline sands, recultivated over 40 years ago, has increased compared to the pure sand 14 times on the average.

The invertebrate's complex of sands, recultivated over 40 years ago a poor variant of mesofauna of taiga podzols of the Kola North with lower taxonomic and trophic diversity and zoomass was represent. Half a century after the carrying out of recultivation of the dormant tailing dump, in newly formed soils there was no formation of zoocenoses, characteristic for zonal podzols of the Kola North.

## 5. CONCLUSIONS

A primary soil-forming process is taking place on ore concentration wastes of apatite-nepheline industry, whose biological recultivation was carried out 40 years ago. The manifestation of the soil-forming process in nepheline sands can be characterized by the following indications: 1) formation of a thin litter with the content of organic carbon at the level of 8-12%; 2) accumulation of humus substances in the mineral sub-litter horizon to the depth of 5 sm as a result of humification of root litter; 3) a distinct change of the response of the medium of the top part (to the depth of 20 sm) of mineral thickness of sands.

Microorganisms are biocatalysts of processes of primary soil formation and one of the primary factors, which determine the specificity of this process. In the process of formation of newly formed soils on nepheline-bearing industrial wastes, there occurs a change in the structure of prokaryotic complex of microbial communities, which originally essentially differed from the prokaryotic complex of zonal soils on moraine sediments. The share of gram-negative bacteria increases in them, while the share of actinobacteria decreases, including streptomycetes. In the formation of pioneer complexes of micromycetes the process of restoration of the vegetation cover, determining their number and diversity, is the most essential, while distinctions in mineralogical and chemical composition of sands, which are parent rock, are not so significant. The recultivation of nepheline sands, carried out 40 years ago, provided prerequisites for the formation of complexes of bacteria and micromycetes, typical for regional podzolic soils.

Common features of invertebrate's complexes in nepheline sands with different period of exposition were the low species diversity and the high level of the number of invertebrates; their colonization by small-size and short-living representatives of micro- and mesofauna; dependence of the succession of pioneer groups of microarthropods on the succession of bacteria and fungi.

The generated ecosystem as a result of biological recultivation and development of the vegetation cover on the surface of nepheline sands represents a natural model of a man-induced formation that underwent a long evolution from barren sands, scarcely occupied only by microorganisms, to complex biogeocenoses, which include the vegetation cover of various structures and the newly formed soil. According to modern classification, soils, generated on reclaimed tailings of apatite industry, can be referred to the grey- humus (cespitose) type with AY-C profile of department of organo-accumulative soils of the post-lithogenic soils's stem [15].

## 6. ACKNOWLEDGEMENTS

We are grateful to N. Mozgova, N. Voronina, L. Baskova, N. Korobeynikova and E. Lebedeva for help in analytic work. This work was sup-

ported by Program "Biodiversity" of Presidium of Russian Academy of Sciences.

## REFERENCES

[1] Ginzburg, K.E. (1981) Phosphorus of basic types of soils of the USSR. Nauka, Moscow.

[2] Pereverzev, V.N., Koshleva, E.A. and Churikov, A.M. (1992) Phosphorus in podzolic soils of the Kola Peninsula. Publishing House of the Kola Science Centre RAS, Apatity.

[3] Pereverzev, V.N. (1993) Cultural soil formation in the Far North. Publishing House of the Kola Science Centre of the RAS, Apatity.

[4] Pereverzev, V.N., Korobeynikova, N.M., Dyakova, T.A. and Yanchenko, I.V. (2007) Agrochemical properties and fertility of soils, generated on tailing dumps of apatite industry after their reclamation. *Agrochemistry*, **1**, 5-12.

[5] Pereverzev, V.N. and Podlesnaya, N.I. (1986) Biological reclamation industrial dumps in the Far North. Publishing House of the Kola Branch of AS of the USSR, Apatity.

[6] Kapelkina, L.P. and Kazakov, L.A. (1989) Wood reclamation of damaged land in the Polar region. *Lesnoye Khozaistvo*, **2**, 27-29.

[7] Evdokimova, G.A., Zenkova, I.V., Mozgova, N.P. and Pereverzev, V.N. (2005) Soil and soil biota in the conditions of fluorine pollution. Publishing House of the Kola Science Centre RAS, Apatity.

[8] Dospekhov, B.A. (1985) The technique of field experience. 5th Edition, Agropromizdat, Moscow.

[9] Gershenkop, A.S., Evdokimova, G. A., Voronina, N.V. and Kreimer, L.L. (2005) Influence of bacterial component of the recycling water on the flotation of not sulphidic ores—by the example of "Apatit". *Inzhenernaya Ekologiya*, **3**, 51-61.

[10] Evdokimova, G.A. and Mozgova, N.P. (2001) Microorganisms of tundra and forest podzols of the Kola North. Publishing House of the Kola Science Centre of the RAS, Apatity.

[11] Evdokimova, G.A. and Naumenko, A.F. (2002) Microorganisms of underground mining workings of Northern Fennoscandia. *Geoecologya. Inzhenernaya Ekologiya. Gydrogeologya. Geokriologya*, **3**, 237-242.

[12] Evdokimova, G.A., Zenkova, I.V. and Pereverzev, V.N. (2002) Biodynamics of processes of organic substance transformation in soils of Northern Fennoscandia. Publishing House of the Kola Science Centre RAS, Apatity.

[13] Evdokimova, G.A., Zenkova, I.V., Mozgova, N.P. and Pereverzev, V.N. (2004) Interactions of soil microorganisms and invertebrate animals at the transformation of plant residues in soils of Northern Fennoscandia. *Soil Science*, **10**, 1199-1210.

[14] Zenkova, I.V. (2002) Succession changes in communities of invertebrate animals in the course of leaf litter decomposing in the zone of influence of copper-nickel companies. *Ecology of Northern Territories of Russia. Problems, Forecast of The Situation, Ways of Development, Solutions*, **2**, 371-375.

[15] Dobrovolsky, G.V. (2004) Classification and diagnostics of soils in Russia. Oykumena, Smolensk.

**Vol.2, No.6, 612-625 (2010)**
doi:10.4236/ns.2010.26076

**Natural Science**

# Application of variational iteration method and electron transfer mediator/catalyst composites in modified electrodes

**Alagu Eswari, Lakshmanan Rajendran***

The Madura College, Madurai, India; *Corresponding Author: raj_sms@rediffmail.com

Received 20 February 2010; revised 23 March 2010; accepted 23 April 2010.

## ABSTRACT

**The nonlinear coupled system of diffusion equations are solved analytically for the transport and kinetics of electrons and reactant in the layer of a modified electrode. Analytical expressions of concentrations of substrate and mediator are presented using He's variational iteration method. The approximate expression of current for microheterogeneous catalysis at isonomer or redox polymer modified electrodes is also obtained. The results of the available limiting cases are compared with our results and are found to be in good agreement.**

**Keywords:** Variational Methods; Nonlinear Boundary Value Problems; Simulation; Reaction-Diffusion Equations; Mathematical Modeling

## 1. INTRODUCTION

Recently the electrocatalytic activity of polymer modified electrodes has been the subject of considerable study of many researchers. Generally most systems are used to require an efficient electron transfer mediator in addition to displaying good electrocatalytic activity. We can find many features for electrocatalysis in the use of microscopic particles of metals or metal oxides dispersed within polymeric films deposited on electrode surfaces. This dispersion of catalytic materials offers important catalytic advantages. We are interested in the design of such microheterogeneous systems for efficient electrocatalysis.

A modified electrode differs from an ordinary electrode by having a thin film of some coated material which prevent direct contact between the metal surface and the bulk electrolyte. Electrochemical reactions of species in solution take place through that thin film. Instead of a direct electron transfer between the Fermi level of the metal and the ion in solution, the electron transfer is 'mediated' by the redox groups present in the thin layer. A large number of different modified electrodes have been made and certain systems have received more attention than others.

Lyons, McCormack, and Bartlett [1] presented an analytical model which quantified the transport and kinetics in conducting polymer modified electrodes containing a homogeneous distribution of spherical microparticulate catalysts. In their paper Lyons, McCormack, and Bartlett [1] obtained the analytical expressions of the substrate and mediator concentrations for the different values of parameters. Lyons and Bartlett [2] also presented the analytical expressions of substrate and mediator concentrations only for limiting values of dimensionless parameters. The transport and kinetics of reactions in chemically modified electrodes have been analyzed previously by Lyons and co-workers and approximate analytical solutions are available [3-7] .

In this paper we analyze application of catalyst composites in modified electrodes. To date many researches have been done on the application of modified electrodes. To my knowledge no rigorous analytical solutions for substrate and mediator concentrations have been reported in that application. Hence the main objective of this paper is to derive the analytical expressions of concentrations of substrate and mediator for all values of parameters using variational iteration method. The approximate expression of current for microheterogeneous catalysis at isonomer or redox polymer modified electrodes is also obtained for all values of dimensionless parameters $\beta$ and $\gamma$ (these parameters are defined in the (7)).

## 2. MATHEMATICAL FORMULATION OF THE PROBLEM AND ANALYSIS

Steady state boundary value problems described the tra-

nsport and kinetics within the film can be written in dimensionless form as follows [2]:

$$D_A \frac{d^2a}{dx^2} - \frac{4\pi r_0^{\,2} N D_A (k'_{A,0} k'_{S,0})^{1/2} (as)^{1/2}}{[D_A + r_0 (k'_{A,0} k'_{S,0})^{1/2} (s/a)^{1/2}]} = 0 \quad (1)$$

$$D_S \frac{d^2s}{dx^2} - \frac{4\pi r_0^{\,2} N D_S (k'_{A,0} k'_{S,0})^{1/2} (as)^{1/2}}{[D_S + r_0 (k'_{A,0} k'_{S,0})^{1/2} (a/s)^{1/2}]} = 0 \quad (2)$$

where $k'_A$ *and* $k'_S$ are electrochemical rate constant, $r_0$ is the radius of an electrode, $D_S$ is the diffusion co-efficient for the substrate concentration, $D_A$ is the diffusion co-efficient for the mediator, $N$ is the number of particles per unit volume, $a$ is the concentration of the mediator and $s$ is the concentration of the substrate. These coupled non-linear differential equations have to be solved by applying the following boundary conditions:

$$x = 0,\ ds/dx = 0 \ \textit{and} \quad a = a^{\infty} \quad (3)$$

and

$$x = L, \quad s = s^{\infty} \ \text{and}\ da/dx = 0 \quad (4)$$

The following dimensionless parameters for substrate concentration $u$, mediator concentration $v$, and distance $X$ are introduced:

$$u = s/s^{\infty}, v = a/a^{\infty}, X = x/X_K \quad (5)$$

where $X_k$ denotes a reaction layer thickness defined by the relation:

$$X_k = \left(\frac{1}{4\pi r_0 N}\right)^{1/2} \quad (6)$$

We also introduce the dimensionless parameters $\beta$ and $\gamma$ defined according to the following relations:

$$\beta = (a^{\infty} D_A / s^{\infty} D_S)^{1/2}, \quad \gamma = r_0 (k'_{A,0}\, k'_{S,0} / D_A D_S)^{1/2} \quad (7)$$

Now the **Eq.1** and **Eq.2** reduce to the following dimensionless form:

$$\frac{d^2u}{dx^2} - \frac{\gamma\beta (uv)^{1/2}}{1 + \gamma\beta (v/u)^{1/2}} = 0 \quad (8)$$

for the substrate, and

$$\frac{d^2v}{dx^2} - \frac{\gamma (uv)^{1/2}}{\beta + \gamma (u/v)^{1/2}} = 0 \quad (9)$$

for the mediator respectively. Here we can assume that the reaction layer thickness $X_K = L$. Now the boundary conditions may be expressed as follows:

$$x = 0\ ,\ v = 1 \ \textit{and} \ \ du/dx = 0 \quad (10)$$

$$x = 1, \ \ dv/dx = 0 \ \ \text{and} \ \ u = 1 \quad (11)$$

The flux $j$ is given by

$$j = -D_A (da/dx)_{x=0} = D_S (ds/dx)_{x=L} \quad (12)$$

or in non-dimensional form:

$$\begin{aligned} j &= -D_A a^{\infty} (4\pi r_0 N)^{1/2} (dv/dx)_{x=0} \\ &= D_S s^{\infty} (4\pi r_0 N)^{1/2} (du/dx)_{x=1} \end{aligned} \quad (13)$$

The required expression of the normalized current is

$$I = \frac{j}{D_S\, s^{\infty} (4\pi r_0 N)^{1/2}} = (du/dx)_{x=1} \quad (14)$$

or

$$I = \frac{j}{D_A\, a^{\infty} (4\pi r_0 N)^{1/2}} = -(dv/dx)_{x=0} \quad (15)$$

## 2.1. Case by Case Transport and Kinetics Analysis

**Case-1: Transport and Kinetics of the Substrate within the Layer**

We consider initially the master **Eq.8** describing the transport and kinetics of the substrate in the layer when $\gamma\beta << 1$. **Eq.8** can be written as

$$\frac{d^2u}{dx^2} - (\gamma\beta) u^{1/2} v^{1/2} + (\gamma^2\beta^2) v = 0 \quad (16)$$

Similarly when $\gamma\beta >> 1 \quad (or) \quad 1/\gamma\beta << 1$, the (8) reduces to

$$\frac{d^2u}{dx^2} - u + (1/\gamma\beta) u^{1/2} v^{-1/2} = 0 \quad (17)$$

The above equations are non-linear and only approximation solutions may be found. Using variational iteration method (Appendix-A), we obtain the concentration of the substrate (by solving the **Eq.16**)

$$\begin{aligned} u(x) = (1+a)^2 &- p(0.033a^2x^6 - 0.1a^2x^5 \\ &- 0.083ax^4 - 0.083a^2x^4 + 0.083ax^4 \\ &+ 0.33ax^3 + 0.33a^2x^3 - 0.5x^2 - 0.5ax^2) \\ &+ p^2(-0.033a^2x^6 + 0.2a^2x^5 - 0.167ax^4 \\ &- 0.33a^2x^4 + 0.667ax^3 - 0.5x^2) \end{aligned} \quad (18)$$

when $\gamma\beta = p << 1$. Using the boundary condition (11) we obtain the following relation between $p$ and $a$.

$$\begin{aligned} a = &-0.25p + 0.239p^2 + 0.108p^3 \\ &- 0.076p^4 - 0.075p^5 \end{aligned} \quad (19)$$

From the above relation we obtain the values of $a$ for any given values of $p$ << 1. The numerical values of $a$ for some given values of $p$ are given in **Table 1**. When $x$ is small, concentration of the substrate (when $\gamma\beta = p << 1$) **Eq.18** becomes

$$u(x) \approx (1+a)^2 + 0.5p(1+a-p)x^2 \quad (20)$$

Also from the above **Eq.20** $u \approx 1$ when $\gamma\beta = p = 0$ ($a = 0$) and $x = 1$. When $\gamma\beta \to 0$, the **Eq.8** or **Eq.16** becomes $d^2u/dx^2 = 0$. The solution of this equation using the boundary conditions (10) and (11) becomes $u = 1$. This result is exactly equal to our result when $\gamma\beta \to 0$. Similarly, the concentration of the substrate becomes (by solving the **Eq.17**

$$\begin{aligned} u(x) = (1+a)^2 &+ (0.5 + 0.5a^2 + a - 0.5r - 0.5ar)x^2 \\ &- 0.333ar(1+a)x^3 + (-0.1667a^2 - 0.1667a \\ &+ 0.1667ar - 0.25a^2r - 0.333a^3r)x^4 + (0.3a^2r \\ &- 0.2a^3r - 0.4a^4r)x^5 + 0.033a^2x^6 \end{aligned} \quad (21)$$

when $\gamma\beta >> 1$ *(or)* $1/\gamma\beta = r << 1$. Using the boundary condition (11), we obtain the following relation between $r$ and $a$.

$$\begin{aligned} &0.4a^4r + 0.533a^3r - (1.3663 - 0.283r)a^2 \\ &- (2.8333 - 0.6663r)a - 0.5(1-r) = 0 \end{aligned} \quad (22)$$

From the above relation we obtain the values of $a$ for any given values of $r$ << 1. We can find the numerical values of $a$ through some specific values of $r$ as shown in **Table 1**. When $x$ is small, **Eq.21** becomes

$$u(x) \approx (1+a)\,[1 + a + 0.5(1+a-r)x^2 - 0.333arx^3] \quad (23)$$

Also from the above **Eq.23**, $u \approx 1$ when $1/\gamma\beta = r = 0$ ($a$ = -1.878 or -0.195) and $x = 1$. When $\gamma\beta \to \infty$, the **Eq.8** or **Eq.17** becomes $d^2u/dx^2 - u = 0$. The solution of this equation using the boundary conditions (10) and (11) becomes $u = 1$. This result is exactly equal to our result when $\gamma\beta \to \infty$. These approximants for the concentration of the substrate **Eq.20**, **Eq.23** are the simplest closed- form of analytical approximation for $\gamma\beta << 1$ and $\gamma\beta >> 1$.

**Case 2: Transport and Kinetics of the Mediator Within the Layer**

We consider the master (9) describing the transport and kinetics of the mediator with in the layer when $\gamma/\beta << 1$. Now the (9) takes the form

$$\frac{d^2v}{dx^2} - (\gamma/\beta)u^{1/2}v^{1/2} + (\gamma^2/\beta^2)u = 0 \quad (24)$$

where as when $\gamma/\beta >> 1$ (9) becomes

$$\frac{d^2v}{dx^2} - v + (\beta/\gamma)v^{3/2}u^{-1/2} = 0 \quad (25)$$

**Figure 1** shows our schematic representation of the differential equations describing the transport and kinetics in a microheterogeneous system. Each of these expressions represents the approximations to the set of master equations outlined in the (8) and (9). Concentration of the mediator using variational iteration method (Appendix -A) becomes (by solving the **Eq.24**

$$\begin{aligned} v(x) = 1 &- 4ax - q(0.033a^2x^6 - 0.1a^2x^5 \\ &- 0.083ax^6 - 0.083a^2x^4 + 0.083ax^4 \\ &+ 0.33ax^3 + 0.33a^2x^3 - 0.5x^2 - 0.5ax^2) \\ &+ q^2(-0.033a^2x^6 + 0.167ax^4 + 0.167a^2x^4 \\ &- 0.5x^2 - ax^2 - 0.5a^2x^2) \end{aligned} \quad (26)$$

when $\gamma/\beta = q << 1$. Using the boundary condition (11) we obtain the following relation between $a$ and $q$

$$a = 0.25q - 0.239q^2 - 0.099q^3 + 0.078q^4 \quad (27)$$

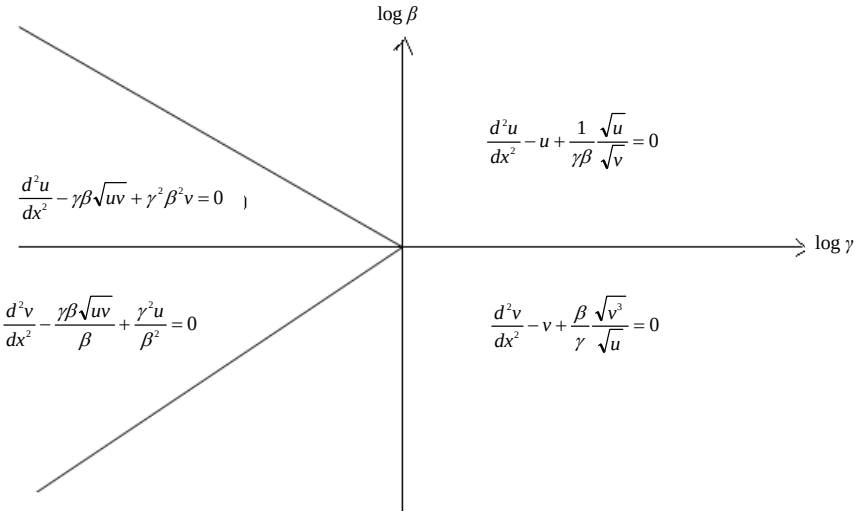



**Figure 1.** Our schematic representation of the differential equations describing the transport and kinetics in a microheterogeneous system. Each of these expressions represents approximations to the set of master equations outlined in the **Eqs.8** and **9**.

The numerical values of $a$ for some given values of $q$ are given in **Table 1**. When $x$ is small, **Eq.26** becomes

$$v(x) \approx 1 - 4ax + 0.5q(1 + a - q - 2aq - a^2 q)x^2 \quad (28)$$

Also from the above **Eq.28**, $v \approx 1$ when $\gamma/\beta = q = 0$ ($a$ = 0) and $x = 0$. When $\gamma/\beta \to 0$, the **Eq.9** or **Eq.24** becomes $d^2v/dx^2 = 0$. The solution of this equation using the boundary conditions (10) and (11) becomes $v = 1$. This result is exactly equal to our result when $\gamma/\beta \to 0$. Similarly, the concentration of the mediator becomes (by solving the **Eq.25**)

$$\begin{aligned} v(x) = 1 - 4ax &+ [0.5 - 0.5l(1 - a + a^2 - a^3)]x^2 \\ &+ [-0.667a + 1.0002al(1 - a + a^2 - a^3)]x^3 \\ &+ [0.333a^2 + 0.1667a - 0.083l(4a + 7a^2 \\ &- 6a^3 + 9a^4 - 12a^5)]x^4 + [-0.2a^2 - 0.05l \\ &(16a^3 - 22a^4 + 4a^5 + 8a^6 - 18a^2)]x^5 \\ &+ 0.033a^2 x^6 \end{aligned} \quad (29)$$

when $\gamma/\beta >> 1$ *(or)* $\beta/\gamma = l << 1$. Using the boundary condition (11) we obtain the following relation between $l$ and $a$

$$\begin{aligned} &2a^6 l - 2.996a^5 l + 0.497a^4 l \\ &-1.998a^3 l + (1.831l - 0.533)a^2 \\ &-(2.668l - 5.333)a - 1 + l = 0 \end{aligned} \quad (30)$$

We can find the numerical values of $a$ through some specific values of $l$ as shown in **Table 1.** When $x$ is small, **Eq.29** becomes

$$\begin{aligned} v(x) \approx 1 - 4ax &+ 0.5[1 - l(1 - a + a^2 - a^3)]x^2 \\ &+ [-0.667a + al(1 - a + a^2 - a^3)]x^3 \end{aligned} \quad (31)$$

Also from the above **Eq.31** $v \approx 1$ when $\beta/\gamma = l = 0$ ($a$ = 0.191 or 9.814) and $x = 0$. When $\beta/\gamma \to \infty$, the (9) or **Eq.25** becomes $d^2v/dx^2 - v = 0$. The solution of this equation using the boundary conditions (10) and (11) becomes $v = 1$. This result is exactly equal to our result when $\beta/\gamma \to \infty$. **Eqs.28** and **31** represent the approximate new analytical expression of the concentration of the mediator when $\gamma/\beta = q << 1$ and $\beta/\gamma = l << 1$. Concentration of substrate and mediator are summarized in **Table 2** and **Table 3.** Using **Eqs.14** and **15** the normalized current $I$ for various cases is given by

$$I = (1 - 0.367a^2)p + (-0.533a^2 + 1.333a - 1)p^2 \quad \text{for } p = \gamma\beta << 1 \quad (32)$$

$$I = 4a \quad \text{for } q = \gamma/\beta << 1 \quad (33)$$

$$\begin{aligned} I = 1 - r + 0.533a^2 + 1.3332a - 1.3322ar - 0.499a^2 r \\ - 2.332a^3 r - 2a^4 r \end{aligned} \quad \text{for } r = 1/\gamma\beta << 1 \quad (34)$$

$$I = 4a \quad \text{for } l = \beta/\gamma << 1 \quad (35)$$

The expression of the current is summarized in **Table 4** and **Table 5**.

**Table 1.** Numerical values of $a$ for various values of $p, r, q, l$ calculated using **Eqs.19**, **22**, **27** and **30**.

| Values of $p, r, q$ and $l$ | Values of $a$ | | | |
|---|---|---|---|---|
| | **Eq.19** | **Eq.22** | **Eq.27** | **Eq.30** |
| 0 | 0 | -0.1950 | 0 | 0.1910 |
| 0.01 | -0.00246 | -0.1931 | 0.0025 | 0.1901 |
| 0.25 | -0.0462 | -0.1520 | 0.0463 | 0.1616 |
| 0.5 | -0.0587 | -0.1056 | 0.0577 | 0.1240 |

**Table 2**. Concentration of substrate $u(x)$ when $\gamma\beta \le 1$ and $\gamma\beta \ge 1$.

| s.no | Conditions | $u(x)$ | | Figures |
|---|---|---|---|---|
| | | This work | Lyons and Bartlett [2] | |
| 1. | $\gamma\beta << 1$ (or) $\gamma\beta = p \le 1$ | $u(x) = (1+a)^2 + 0.5p(1+a-p)x^2$ (20) where $a = -0.25p + 0.239p^2 + 0.108p^3 - 0.076p^4 - 0.075p^5$ | $u(x) = \left[1 - 0.2887\sqrt{p}(1-x)\right]^4$ (40) | Figure 3 $\gamma\beta$ = 0.01,0.25, 0.5 Figure 11 $\gamma\beta$ = 0.5 |
| 2. | $\gamma\beta >> 1$ (or) $1/\gamma\beta = r \le 1$ | $u(x) = (1+a)\,[1 + a + 0.5(1+a-r)x^2 - 0.333arx^3\,]$ (23) where $0.4a^4 r + 0.533a^3 r - (1.3663 - 0.283r)a^2 - (2.8333 - 0.6663r)a - 0.5(1-r) = 0$ | $u(x) = \cosh(x)\sec h$ (41) | Figure 4 $\gamma\beta$ = 2,4,100 Figure 12 $\gamma\beta$ = 4 |

**Table 3.** Concentration of mediator $v(x)$ when $\gamma/\beta \le 1$ and $\gamma/\beta \ge 1$.

| s.no | Conditions | $v(x)$ This work | | $v(x)$ Lyons and Bartlett [2] | Figures |
|---|---|---|---|---|---|
| 1. | $\gamma/\beta << 1$ (or) $\gamma/\beta = q \le 1$ | $v(x) \approx 1 - 4ax + 0.5q(1 + a - q - 2aq - a^2 q)x^2$ where $a = 0.25q - 0.239q^2 - 0.099q^3 + 0.078q^4$ | (28) | -------- | Figure 5 $\gamma/\beta$ = 0.01,0.25,0.5 Figure 13 $\gamma/\beta$ = 0.5 |
| 2. | $\gamma/\beta >> 1$ (or) $\beta/\gamma = l \le 1$ | $v(x) \approx 1 - 4ax + 0.5[1 - l(1 - a + a^2 - a^3)]x^2 + [-0.667a + al(1 - a + a^2 - a^3)]x^3$ where $2a^6 l - 2.996a^5 l + 0.497a^4 l - 1.998a^3 l + (1.831l - 0.533)a^2 - (2.668l - 5.333)a - 1 + l = 0$ | (31) | $v(x) = \cosh(x) - \tanh\sinh(x)$ (42) | Figure 6 $\gamma/\beta$ = 2,4,100 Figure 14 $\gamma/\beta$ = 2 |

**Table 4**. Current $I$ when $\gamma\beta \le 1$ and $\gamma\beta \ge 1$.

| s.no | Conditions | Current $I$ This work | | Current $I$ Lyons and Bartlett [2] | | Figures |
|---|---|---|---|---|---|---|
| 1. | $\gamma\beta << 1$ (or) $\gamma\beta = p \le 1$ | $I = (1 - 0.367a^2)p + (-0.533a^2 + 1.333a - 1)p^2$ | (32) | $I = \sqrt{p}$ | (43) | Figure 7 Figure 15 |
| 2. | $\gamma\beta >> 1$ (or) $1/\gamma\beta = r \le 1$ | $I = 1 - r + 0.533a^2 + 1.3332a - 1.3322ar - 0.499a^2 r - 2.332a^3 r - 2a^4 r$ | (34) | $I = \tanh$ | (44) | Figure 8 Figure 16 |

**Table 5.** Current $I$ when $\gamma/\beta \le 1$ and $\gamma/\beta \ge 1$.

| s.no | Conditions | Current $I$ This work | | Current $I$ Lyons and Bartlett [2] | | Figures |
|---|---|---|---|---|---|---|
| 1. | $\gamma/\beta << 1$ (or) $\gamma/\beta = q \le 1$ | $I = 4a$ | (33) | $I = \sqrt{q}$ | (45) | Figure 9 Figure 17 |
| 2. | $\gamma/\beta >> 1$ (or) $\beta/\gamma = l \le 1$ | $I = 4a$ | (35) | $I = \tanh$ | (46) | Figure 10 Figure 18 |

## 3. COMPARISON WITH LYONS AND BARTLETT [2] WORK

Lyons and Bartlett [2] takes the (8) in the form

$$d^2u/dx^2 - \gamma\beta(uv)^{1/2} = 0 \qquad (36)$$

when $\gamma\beta << 1$ whereas

$$d^2u/dx^2 - u = 0 \qquad (37)$$

when $\gamma\beta >> 1$. The third term in the **Eqs.16** and **17** is not found in the **Eqs.36** and **37**. When $\gamma << \beta$ (9) takes the form:

$$\frac{d^2v}{dx^2} - \frac{\gamma(uv)^{1/2}}{\beta} = 0 \qquad (38)$$

whereas when $\gamma >> \beta$

$$d^2v/dx^2 - v = 0 \quad (39)$$

The third term in the **Eqs.24** and **25** is not present in the **Eqs.38** and **39**. **Figure 2** shows the schematic representation of the differential equations describing the transport and kinetics in a microheterogeneous system by (Lyons and Bartlett [2]). Each of these expressions represents the approximations to the set of master equations outlined in the **Eqs.8** and **9**. Lyons and Bartlett [2]) obtained the concentration of the substrate as

$$u(x) = \left[1 - 0.2887\sqrt{p}(1-x)\right]^4 \quad \text{if} \quad p = \gamma\beta << 1 \quad (40)$$

$$u(x) = \cosh(x)\sec h \quad \text{if} \quad p = \gamma\beta >> 1 \quad (41)$$

Similarly, mediator concentration as

$$v(x) = \cosh(x) - \tanh\sinh(x) \quad \text{if} \quad l = \beta/\gamma << 1 \quad (42)$$

But a definite solution for mediator concentration is not arrived upon in the case of $q = \gamma/\beta << 1$ . The **Eqs.40** and **41** derived by Lyons and Bartlett [2] satisfy the boundary condition (11), but the **Eq.40** does not satisfy the boundary condition (10). In the same way **Eq.41** is independent of the parameter $1/\gamma\beta$ whereas our **Eqs.20** and **23** satisfy the boundary conditions (10) and (11). Similarly **Eq.42** is independent of the parameter $\beta/\gamma$ whereas our **Eqs.28** and **31** satisfy the boundary conditions (10) and (11). Lyons and Bartlett [2] obtained the corresponding dimensionless current $I$ as follows.

$$I = \sqrt{p} \quad \text{if} \quad p = \gamma\beta << 1 \quad (43)$$

$$I = \tanh \quad \text{if} \quad r = 1/\gamma\beta << 1 \quad (44)$$

$$I = \sqrt{q} \quad \text{if} \quad q = \gamma/\beta << 1 \quad (45)$$

$$I = \tanh \quad \text{if} \quad l = \beta/\gamma << 1 \quad (46)$$

**Eqs.44** and **46** are independent of the parameters $1/\gamma\beta$ and $\beta/\gamma$ whereas our **Eqs.34** and **35** depend on the parameters $1/\gamma\beta$ and $\beta/\gamma$ .

## 4. DISCUSSION OF STEADY STATE PROBLEM

The comparison of concentration of substrate $u(x)$ between the **Eqs.20** and **23** (This work) and **Eqs.40** and **41** (Lyons and Bartlett [2]) are represented in **Figure 3**, **Figure 4** for various values of $\gamma\beta$ . From these **Figures** it is understood that the value of the concentration decreases when $\gamma\beta$ increases. Concentration is slowly increasing when $x \geq 0.5$ for all values of $\gamma\beta$ . Then the concentration of $u(x)$ becomes 1 when $x = 1$ for all values of $\gamma\beta$ . The comparison of concentration mediator $v(x)$ between the **Eqs.28** and **31** (This work) and **Eq.42** (Lyons and Bartlett [2]) are represented in **Figure 5** and **Figure 6** for various values of $\gamma/\beta$ . From these figures, it is deducted that the value of the concentration of $u(x)$ decreases when $\gamma/\beta$ increases. Concentration is slowly decreasing when $x \geq 0.6$ for all values of $\gamma/\beta$ . From these **Figures 3-6**, it is constructed that **Eqs.20**, **23**, **28** and **31** satisfy their boundary conditions (10) and (11).

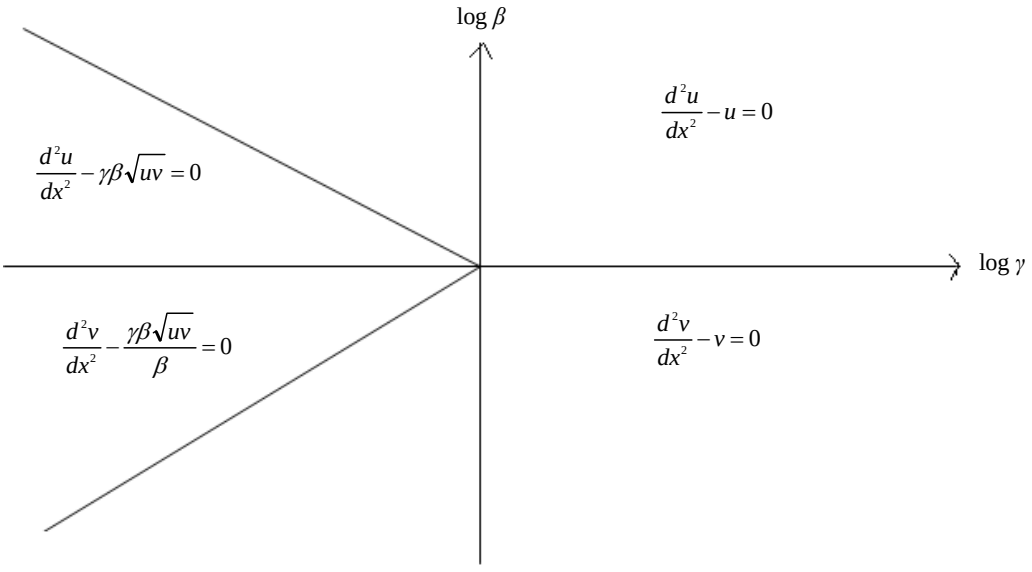



**Figure 2.** Schematic representation of the differential equations describing the transport and kinetics in a microheterogeneous system by Lyons and Bartlett [2]. Each of these expressions represents approximations to the set of master equations outlined in the **Eqs.8** and **9**.

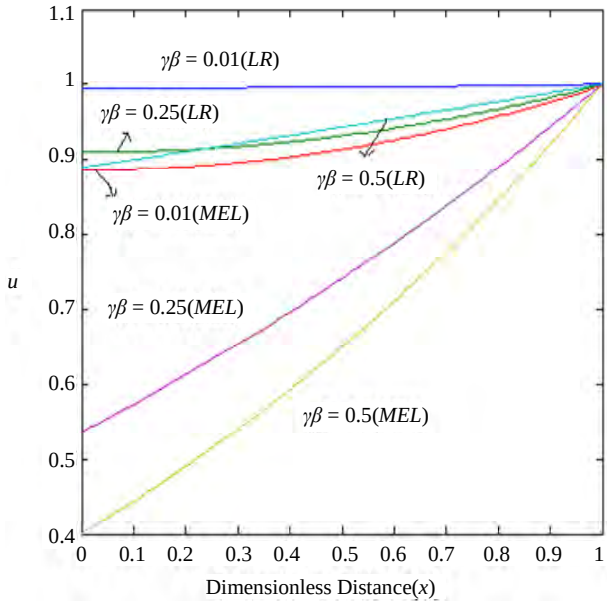



**Figure 3.** Profiles of the dimensionless substrate concentration $u$, for various values of $\gamma\beta$ when $\gamma\beta << 1$. The curve is plotted using **Eqs.20** and **40**.

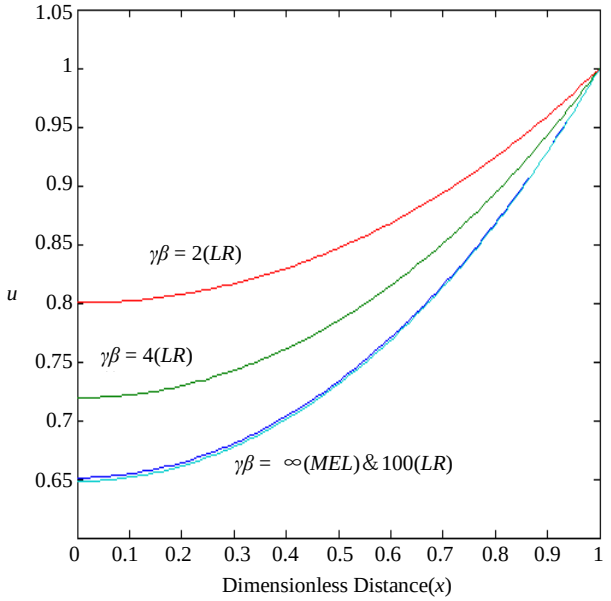



**Figure 4.** Profiles of the dimensionless substrate concentration $u$, for various values of $\gamma\beta$ when $\gamma\beta >> 1$ .The curve is plotted using **Eqs.23** and **41**.

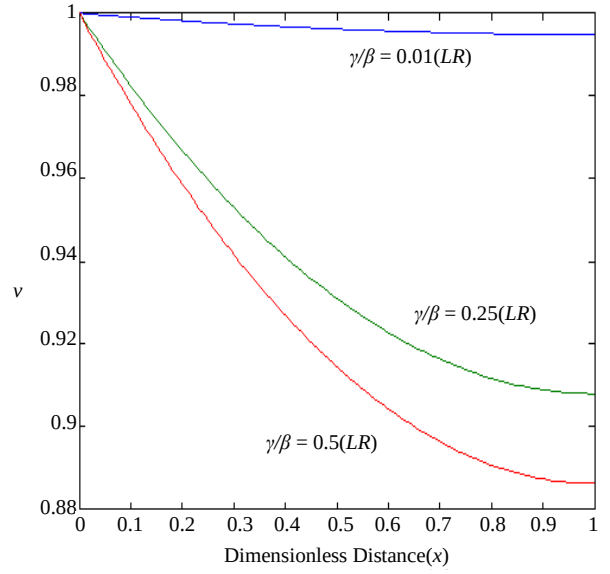



**Figure 5.** Profiles of the dimensionless mediator concentration $v$, for various values of $\gamma/\beta$ when $\gamma/\beta << 1$. The curve is plotted using **Eq.28**.

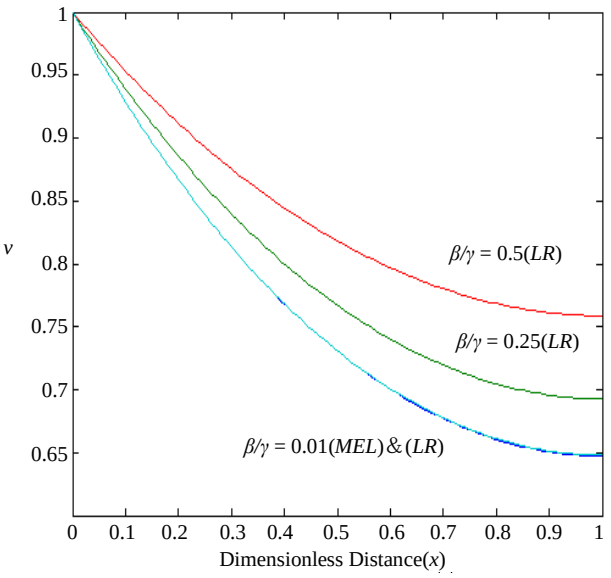



**Figure 6.** Profiles of the dimensionless mediator concentration $v$, for various values of $\gamma/\beta$ when $\gamma/\beta >> 1$. The curve is plotted using **Eqs.31** and **42**.

**Figure 7** shows the comparison of dimensionless current $I$ between the **Eqs.32** and **43** for various values of $\gamma\beta$ when $\gamma\beta << 1$. **Figure 8** shows the comparison of dimensionless current $I$ between the **Eqs.34** and **44** for various values of $\gamma\beta$ when $\gamma\beta >> 1$. **Figure 9** shows the comparison of dimensionless current $I$ between the **Eqs.33** and **45** for various values of $\gamma/\beta$ when $\gamma/\beta << 1$. **Figure 10** shows the comparison of dimensionless current $I$ between the **Eqs.35** and **46** for various values of $\gamma/\beta$ when $\gamma/\beta >> 1$. In all case diagrams as shown in figures, there is a vast variation in the current curves.

**Figures 11** and **12** show the comparison of our di-

mensionless concentration $u$ evaluated using **Eqs.20** and **23** (This work) together with the simulation results (This work) and **Eqs.40** and **41** (Lyons and Bartlett [2]) for the case of $\gamma\beta << 1$ and $\gamma\beta >> 1$. **Figures 13** and **14** indicate the comparison of our dimensionless concentration $v$ calculated using **Eqs.28** and **31** (This work) together with the simulation results (This work) and **Eq.42** (Lyons and Bartlett [2]) for the case of $\gamma/\beta << 1$ and $\gamma/\beta >> 1$. In all cases, there is a good match between our analytical and simulation results.

**Figures 15** and **16** show the comparison of our dimensionless current $I$ versus $\gamma\beta$ evaluated using **Eqs.32** and **34** (This work) together with the simulation results (This work) and **Eqs.43** and **44** (Lyons and Bartlett [2]) for the case of $\gamma\beta << 1$ and $\gamma\beta >> 1$. **Figures 17** and **18** indicate the comparison of our dimensionless current $I$ versus $\gamma/\beta$ calculated using **Eqs.33** and **35** (This work) together with the simulation results (This work) and **Eqs.45** and **46** (Lyons and Bartlett [2]) for the case of $\gamma/\beta << 1$ and $\gamma/\beta >> 1$. In all cases, there is a good

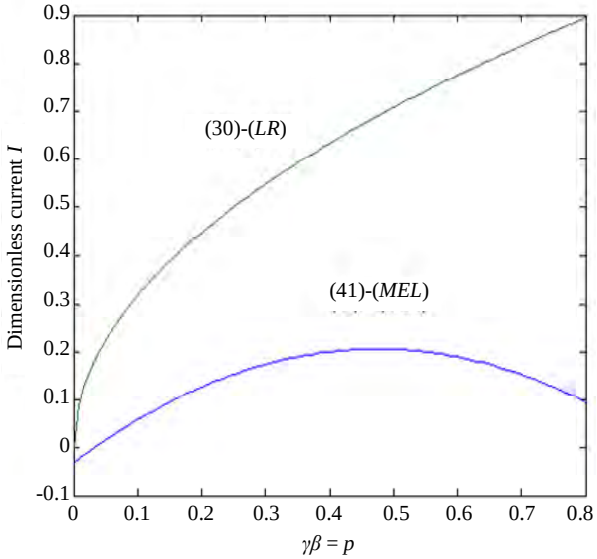



**Figure 7.** Plot of dimensionless current versus $\gamma\beta = p$. Current is calculated in the **Eqs.32** and **43**.

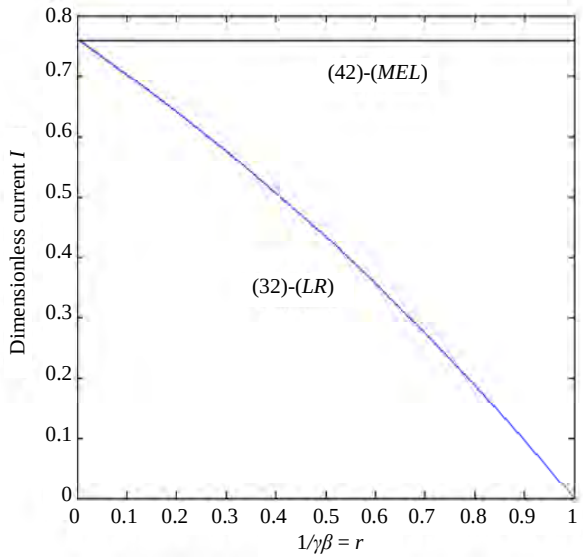



**Figure 8.** Plot of dimensionless current versus $1/\gamma\beta = r$. Current is calculated in the **Eqs.34** and **44**.

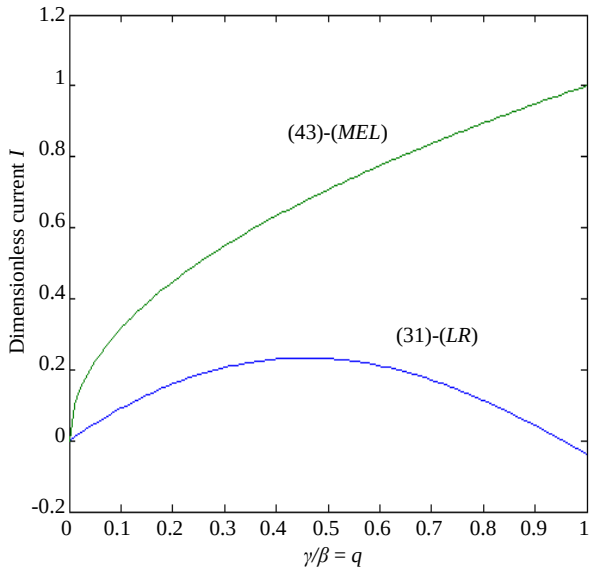



**Figure 9.** Plot of dimensionless current versus $\gamma/\beta = q$. Current is calculated in the **Eqs.33** and **45**.

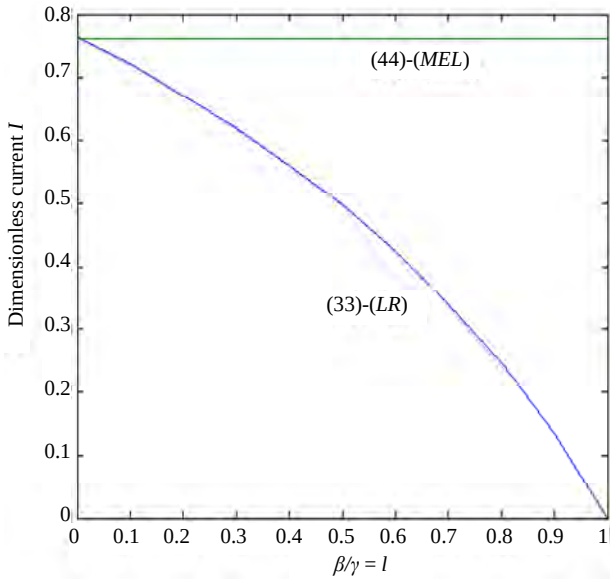



**Figure 10.** Plot of dimensionless current versus $\beta/\gamma = l$. Current is calculated in the **Eqs.35** and **46**.

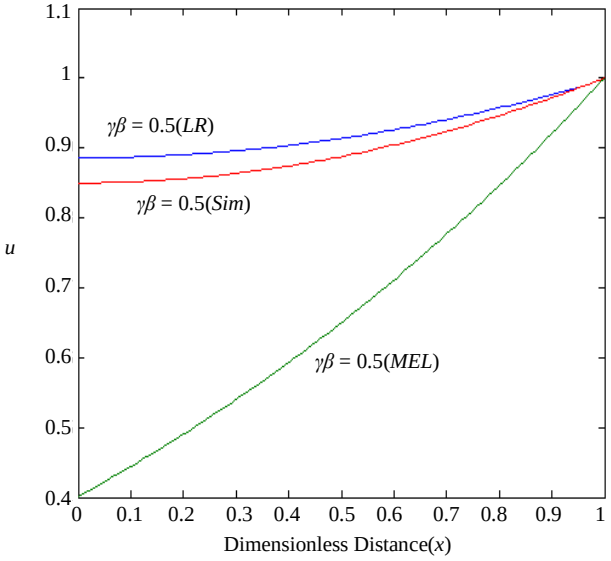



**Figure 11.** Comparison of our dimensionless concentration $u$ using **Eq.20** (This work), **Eq.40** (Lyons and Bartlett [2]) and simulation results using Scilab ( This work). Here the value of $\gamma\beta = 0.5 << 1$.

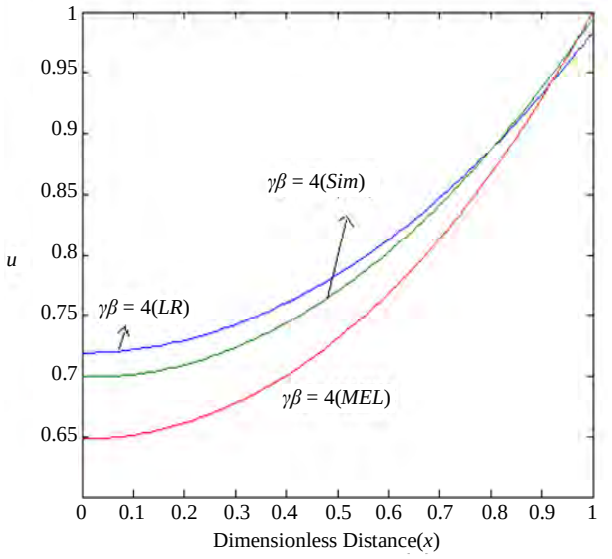



**Figure 12.** Comparison of our dimensionless concentration $u$ using **Eq.23** (This work), **Eq.41** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma\beta = 4 >> 1$.

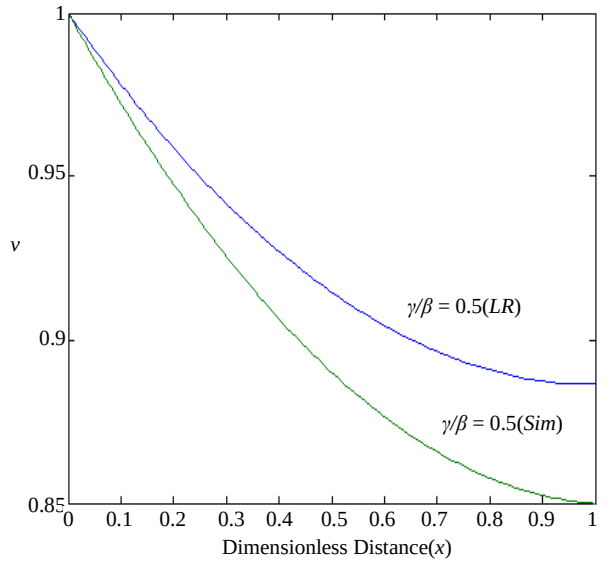



**Figure 13.** Comparison of our dimensionless concentration $v$ using **Eq.28** (This work) and simulation results using Scilab (This work). Here the value of $\gamma/\beta = 0.5 << 1$.

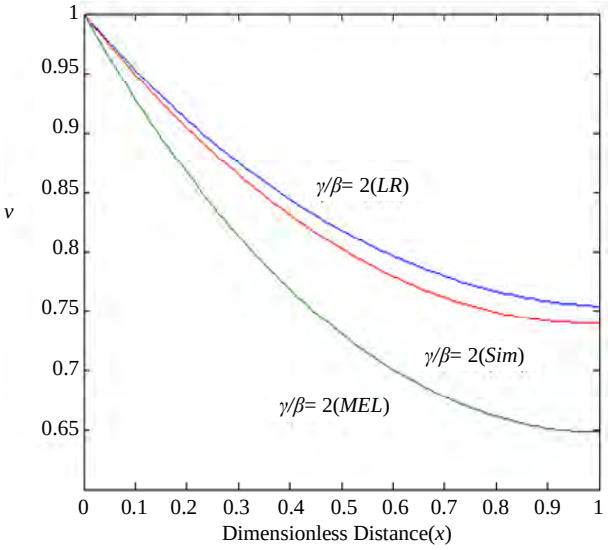



**Figure 14.** Comparison of our dimensionless concentration $v$ using **Eq.31** (This work), **Eq.42** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma/\beta = 2 >> 1$.

match between our analytical and simulation results.

## 5. MATHMATICAL FORMULATION OF THE PROBLEM AND ANALYSIS (NON-STEADY STATE)

The initial boundary value problems which has to be solved for the case of non-steady state can be written in dimensionless form as follows

$$\frac{du}{dt} = \frac{d^2u}{dx^2} - \frac{\gamma\beta(uv)^{1/2}}{1+\gamma\beta(v/u)^{1/2}} \tag{47}$$

for the substrate, and

$$\frac{dv}{dt} = \frac{d^2v}{dx^2} - \frac{\gamma(uv)^{1/2}}{\beta+\gamma(u/v)^{1/2}} \tag{48}$$

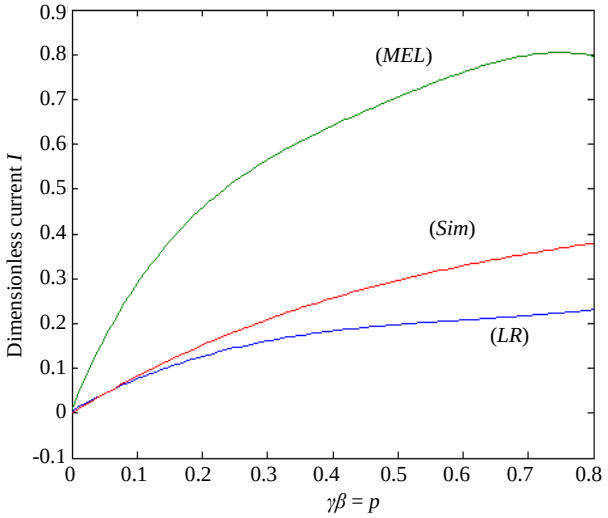



**Figure 15.** Comparison of our dimensionless current $I$ versus $\gamma\beta = p$ using **Eq.32** (This work), **Eq.43** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma\beta << 1$.

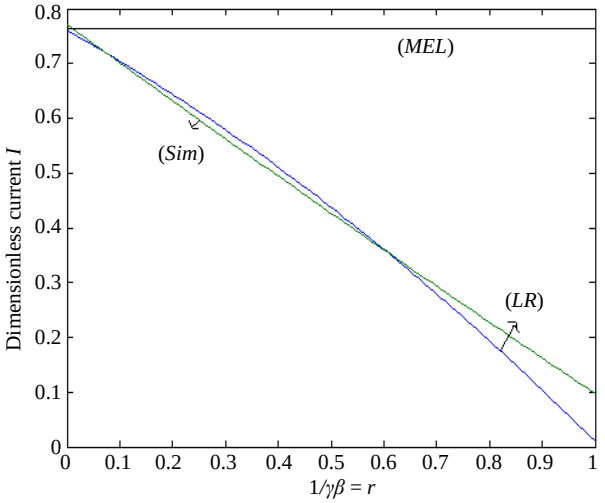



**Figure 16.** Comparison of our dimensionless current $I$ versus $1/\gamma\beta = r$ using **Eq.34** (This work), **Eq.44** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma\beta >> 1$.

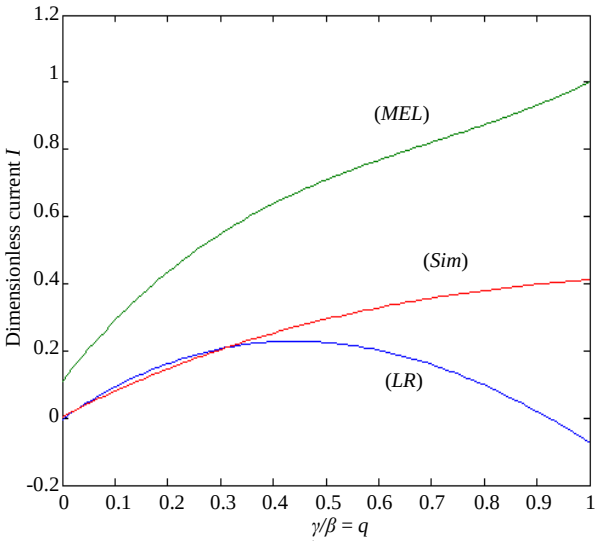



**Figure 17.** Comparison of our dimensionless current $I$ versus $\gamma/\beta = q$ using **Eq.33** (This work), **Eq.45** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma/\beta << 1$.

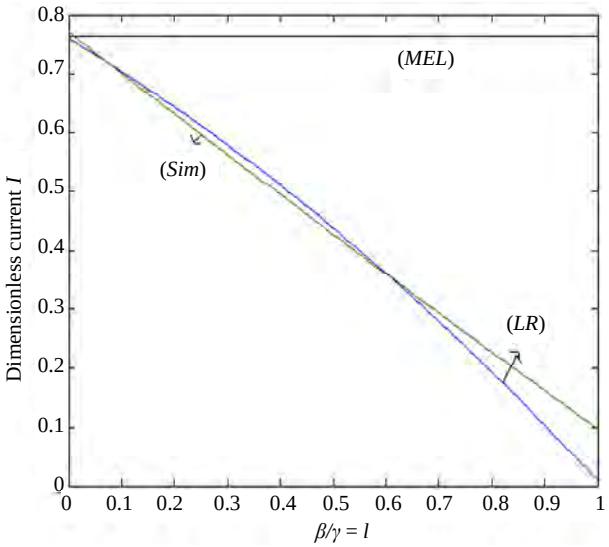



**Figure 18.** Comparison of our dimensionless current $I$ versus $\beta/\gamma = l$ using **Eq.35** (This work), **Eq.46** (Lyons and Bartlett [2]) and simulation results using Scilab (This work). Here the value of $\gamma/\beta >> 1$.

for the mediator respectively. These equations must obey the following initial and boundary conditions

$$t = 0 \ , u = 1 \ and \ v = 0.001 \tag{49}$$

$$x = 0 \ , v = 1 \ \ and \ \ du/dx = 0 \tag{50}$$

$$x = 1, \ \ dv/dx = 0 \ \ \text{and} \ \ u = 1 \tag{51}$$

To my knowledge no rigorous analytical (or) numerical solutions for the transient problems have been reported. Numerical simulation of substrate concentration and mediator concentration can be evaluated using Scilab software (Appendix-B) .

## 6. DISCUSSION OF NON-STEADY STATE PROBLEM

The normalized numerical simulations of three dimensional substrate concentration $u(x)$ is shown in **Figure 19**, **Figure 21** and **Figure 23**. As shown in **Figures 19**,

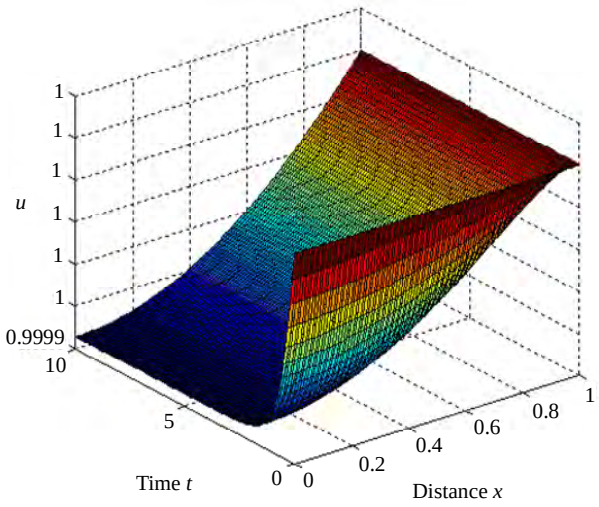



**Figure 19.** The normalized numerical simulation of three dimensional substrate concentration $u(x)$. The plot was constructed using **Eq.47** for $\gamma = 0.01, \beta = 0.01$.

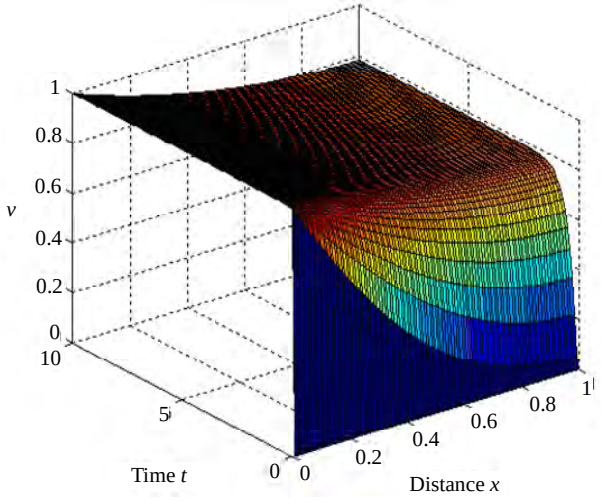



**Figure 20.** The normalized numerical simulation of three dimensional mediator concentration $v(x)$. The plot was constructed using **Eq.48** for $\gamma = 0.01, \beta = 0.01$.

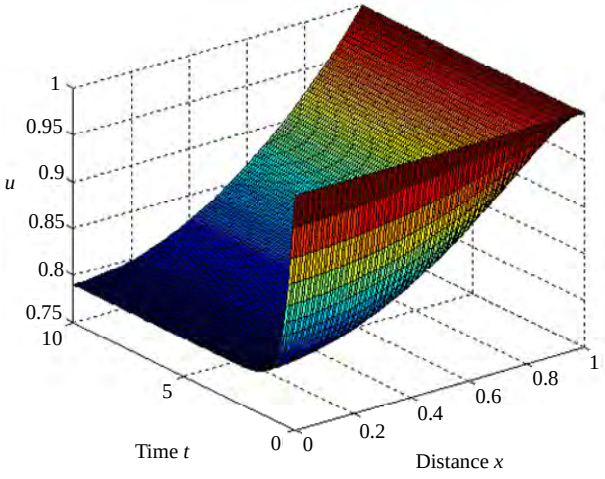



**Figure 21.** The normalized numerical simulation of three dimensional substrate concentration $u(x)$. The plot was constructed using **Eq.47** for $\gamma = 1, \beta = 1$.

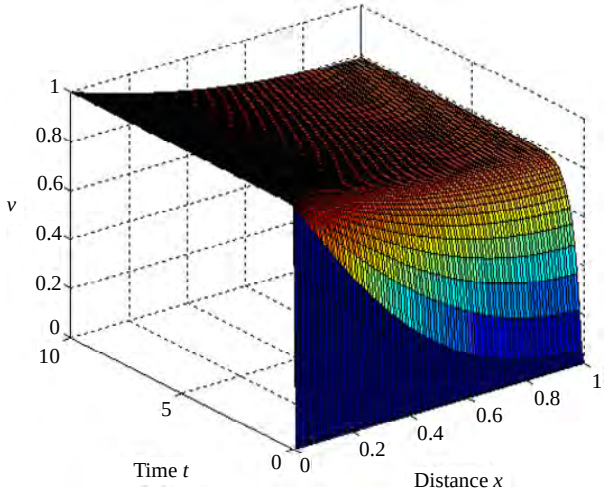



**Figure 22.** The normalized numerical simulation of three dimensional mediator concentration $v(x)$. The plot was constructed using **Eq.48** for $\gamma = 1, \beta = 1$.

**Figure 21** and **Figure 23** give the calculated response curve at different $\gamma$ and $\beta$ values in our diagrams. The time dependent concentration $u(x)$ using **Eq.47** is represented in **Figure 19**, **Figure 21** and **Figure23** for various values of $\gamma$ and $\beta$. Concentration is slowly decreasing when $\gamma$ and $\beta$ is increasing. Then the concentration of $u(x)$ = 1 when $x = 1$ and also for all values of $\gamma, \beta$ and $t$. The normalized numerical simulation of three dimensional mediator concentration $v(x)$ is shown in **Figure 20**, **Figure 22** and **Figure 24**. These figures show the calculated curve at different $\gamma$ and $\beta$ values in our diagrams. The time dependent curve $v(x)$ using **Eq.48** is represented in **Figure 20**, **Figure 22** and **Figure 24** for all values of $\gamma, \beta$ and $t$. $\gamma$ and $\beta$ larger than 1 may be appropriate because the slope of the curves corresponding to $\gamma \geq 1$ and $\beta \geq 1$ are almost identical although the analytical ranges are different. The slope of the curves decreases dramatically and the concentration is identical range when $\gamma$ and $\beta$ is high.

## 7. CONCLUSIONS

We have presented a simple analysis of reaction/diffusion within a conducting polymer film which is deposited on a microelectrode. The transport and kinetics are quantified in terms of a fundamental reaction/diffusion

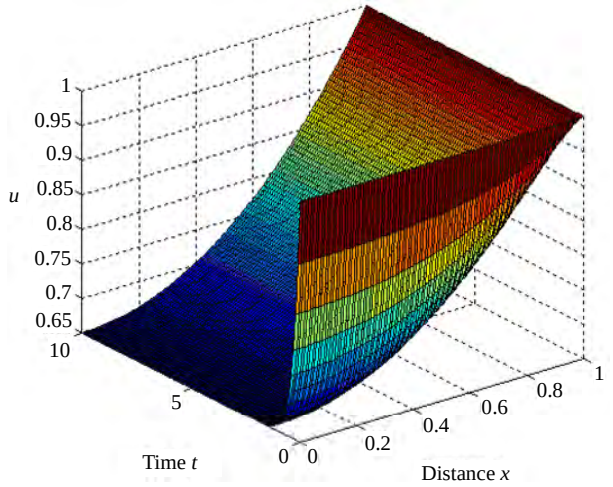



**Figure 23.** The normalized numerical simulation of three dimensional substrate concentration $u(x)$. The plot was constructed using **Eq.47** for $\gamma = 10, \beta = 10$.

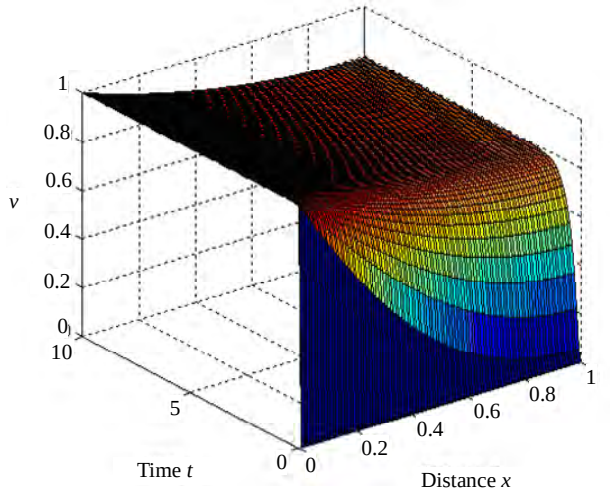



**Figure 24.** The normalized numerical simulation of three dimensional mediator concentration $v(x)$. The plot was constructed using **Eq.48** for $\gamma = 10, \beta = 10$.

parameters $\gamma\beta$, $\gamma/\beta$ and the analytical expression of the substrate concentration and mediator concentration within the polymer film are thus derived. An analytical expression for the steady state current response is also presented. A non linear time independent partial differential equation has been formulated and solved using He's variational iteration method. The primary result of this work is first approximate calculation of substrate concentration and mediator concentration for all values of $\gamma\beta$ and $\gamma/\beta$. It gives good agreement with previous published limiting case results. The extension of the procedure to other two-dimensional and three-dimensional geometries with various complex boundary conditions seems possible.

## 8. ACKNOWLEDGEMENTS

This work was supported by the Department of Science and Technology (DST) Government of India. The authors also thank Mr. M. S. Meenakshisundaram, Secretary, The Madura College Board Dr. T. V. Krishnamoorthy, Principal, and Mr. S. Thiagarajan Head of the Department of Mathematics, The Madura College, Madurai, for their constant encouragement. It is our pleasure to thank the referees for their valuable comments.

## REFERENCES

[1] Lyons, M.E.G., McCormack, D.E. and Bartlett, P.N. (1989) Microheterogeneous catalysis in modified electrodes. *Journal of Electroanalytical Chemistry*, **261(1)**, 51-59.

[2] Lyons, M.E.G. and Bartlett, P.N. (1991) Microheterogeneous catalysis in modified electrodes: Part 2. Electron transfer mediator/catalyst composites. *Journal of Electroanalytical Chemistry*, **316(1-2)**, 1-22.

[3] Harrison, D.J. and Wrighton, M.S. (1984) Catalysis of hydrogen evolution via deposition of palladium onto electrodes modified with an N, N'-Dialkyl-4, 4'-bipyridinium-based polymer: Dependence of rate on palladium coverage. *Journal of Physical Chemistry*, **88(18)**, 3932-3935.

[4] Andrieux, C.P., Dumas-Bouchiat, J.M. and Saveant, J.M. (1982) Catalysis of electrochemical reactions at redox polymer electrodes: Kinetic model for stationary voltammetric techniques. *Journal of Electroanalytical Chemistry*, **131**, 1-35.

[5] Albery, W.J. and Hillman, A.R. (1984) Transport and kinetics in modified electrodes. *Journal of Electroanalytical Chemistry*, **170(1-2)**, 27-49.

[6] Hillman, A.R. and Linford, R.G. (1987) Electrochemical science and technology of polymers. Elsevier, Amsterdam, **1**, 103-291.

[7] Albery, W.J. and Hillman, A.R. (1981) Modified electrodes. *Annual Report Section C of the Royal Society of Chemistry*, **78**, 377-438.

[8] He, J.H. (1999) Variational iteration method–a kind of nonlinear analytical technique: Some examples. *International Journal of Non-Linear Mechanics*, **34(4)**, 699-708.

[9] Momani, S. and Abuasad, S. (2006) Application of He's variational iteration method to Helmholtz equation. *Chaos, Solitons & Fractals*, **27(5)**, 1119-1123.

[10] Abdou, M.A. and Soliman, A.A. (2005) Variational iteration method for solving Burger's and coupled Burger's equations. *Journal of Computational and Applied Mathematics*, **181(2)**, 245-251.

[11] He, J.H. and Wu, X.H. (2006) Construction of solitary solution and compacton-like solution by variational iteration method. *Chaos, Solitons & Fractals*, **29(1)**, 108.

[12] Rajendran, L. and Rahamathunissa, G. (2008) Application of He's variational iteration method in nonlinear boundary value problems in enzyme-substrate reaction diffusion processes: Part 1. The steady-state amperometric response. *Journal of Mathematical Chemistry*, **44(3)**, 849-861.

## APPENDIX A

In this appendix, we derive the general solution of non-linear reaction (8), (9), (10), (11) using He's variational iteration method. To illustrate the basic concepts of variational iteration method (VIM), we consider the following non-linear partial differential equation [8-12]

$$L[u(x)]+N[u(x)]=g(x) \tag{A1}$$

where $L$ is a linear operator, $N$ is a non-linear operator, and $g(x)$ is a given continuous function. According to the variational iteration method, we can construct a correct functional as follows [11]

$$u_{n+1}(x)=u_n(x)+\int_0^x \lambda\left[L[u_n(\tau)]+N[\tilde{u}_n(\tau)]-g(\tau)\right]d\tau \tag{A2}$$

Where $\lambda$ is a general Lagrange multiplier which can be identified optimally via variational theory, $u_n$ is the n$^{th}$ approximate solution, and $\tilde{u}_n$ denotes a restricted variation, i.e., $\delta\tilde{u}_n=0$. In this method, a trial function (an initial solution) is chosen which satisfies given boundary conditions. Using the above variational iteration method we can write the correction functional of **Eqs.16**, **17**, **24** and **25** as follows

$$u_{n+1}(x)=u_n(x)+\int_0^x \lambda\left[u_n''(\xi)+\widetilde{\gamma^2\beta^2 v_n(\xi)}-\widetilde{\gamma\beta u_n^{1/2}(\xi)v_n^{1/2}(\xi)}\right]d\xi \tag{A3}$$

$$v_{n+1}(x)=v_n(x)+\int_0^x \lambda\left[v_n''(\xi)-\widetilde{\frac{\gamma}{\beta}u_n^{1/2}(\xi)v_n^{1/2}(\xi)}+\widetilde{\frac{\gamma^2}{\beta^2}u_n(\xi)}\right]d\xi \tag{A4}$$

$$u_{n+1}(x)=u_n(x)+\int_0^x \lambda\left[u_n''(\xi)-u_n(\xi)+\widetilde{\frac{1}{\gamma\beta}u_n^{1/2}(\xi)v_n^{-1/2}(\xi)}\right]d\xi \tag{A5}$$

$$v_{n+1}(x)=v_n(x)+\int_0^x \lambda\left[v_n''(\xi)-v_n(\xi)+\widetilde{\frac{\beta}{\gamma}\frac{v_n^{3/2}(\xi)}{u_n^{1/2}(\xi)}}\right]d\xi \tag{A6}$$

Taking variation with respect to the independent variables $u_n$ *and* $v_n$, we get

$$\delta u_{n+1}(x)=\delta u_n(x)+\delta\int_0^x \lambda\left[u_n''(\xi)+\widetilde{\gamma^2\beta^2 v_n(\xi)}-\widetilde{\gamma\beta u_n^{1/2}(\xi)v_n^{1/2}(\xi)}\right]d\xi \tag{A7}$$

$$\delta v_{n+1}(x)=\delta v_n(x)+\delta\int_0^x \lambda\left[v_n''(\xi)-\widetilde{\frac{\gamma}{\beta}u_n^{1/2}(\xi)v_n^{1/2}(\xi)}+\widetilde{\frac{\gamma^2}{\beta^2}u_n(\xi)}\right]d\xi \tag{A8}$$

$$\delta u_{n+1}(x)=\delta u_n(x)+\delta\int_0^x \lambda\left[u_n''(\xi)-u_n(\xi)+\widetilde{\frac{1}{\gamma\beta}u_n^{1/2}(\xi)v_n^{-1/2}(\xi)}\right] \tag{A9}$$

$$\delta v_{n+1}(x)=\delta v_n(x)+\delta\int_0^x \lambda\left[v_n''(\xi)-v_n(\xi)+\widetilde{\frac{\beta}{\gamma}\frac{v_n^{3/2}(\xi)}{u_n^{1/2}(\xi)}}\right]d\xi \tag{A10}$$

where $\lambda$ is general Lagrangian multipliers, $u_0$ and $v_0$ are

initial approximations or trial functions $\widetilde{\gamma^2\beta^2 v_n(\xi)}$, $\widetilde{\gamma\beta u_n^{1/2}(\xi)v_n^{1/2}(\xi))}$, $\widetilde{\frac{\gamma}{\beta}u_n^{1/2}(\xi)v_n^{1/2}(\xi)}$, $\widetilde{\frac{\gamma^2}{\beta^2}u_n(\xi)}$ $\widetilde{\frac{1}{\gamma\beta}u_n^{1/2}(\xi)v_n^{-1/2}(\xi)}$, $\widetilde{\frac{\beta}{\gamma}\frac{v_n^{3/2}(\xi)}{u_n^{1/2}(\xi)}}$ and are considered as restricted variations *i.e.* $\delta\tilde{u}_n = 0, \delta\tilde{v}_n = 0$ and $\delta\tilde{u}_n\tilde{v}_n = 0$. Making the above correction functional (A7) to (A10) stationary, noticing that $\delta u_n(0) = 0, \delta v_n(0) = 0$ and $\delta u_n(0)\, v_n(0) = 0$.

$$\delta u_n : 1 - \lambda'(\xi)\big|_{\xi=x} = 0, \delta v_n : 1 - \lambda'(\xi)\big|_{\xi=\tau} = 0 \quad \text{(A11)}$$

$$\delta u'_n : \ \lambda(\xi)\big|_{\xi=x} = 0, \quad \delta v'_n : \lambda(\xi)\big|_{\xi=x} = 0 \quad \text{(A12)}$$

$$\delta u_n : \lambda''(\xi)\big|_{\xi=x} = 0, \quad \delta v_n : \lambda''(\xi)\big|_{\xi=x} = 0 \quad \text{(A13)}$$

The above equations are called Lagrange-Euler equations. By solving the above equations the Lagrange multipliers, can be identified as

$$\lambda(\xi) = (\xi - x) \quad \text{(A14)}$$

substituting the Lagrangian multipliers and $n = 0$ in the iteration formula (A3, A4, A5, A6) we obtain,

$$u_1(x) = u_0(x) + \int_0^x (\xi - x)[u_0''(\xi) - \gamma\beta u_0^{1/2}(\xi)v_0^{1/2}(\xi) + \gamma^2\beta^2 v_0(\xi)]d\xi \quad \text{(A15)}$$

$$u_1(x) = u_0(x) + \int_0^x (\xi - x)[u_0''(\xi) - u_0(\xi) + (1/\gamma\beta)u_0^{1/2}(\xi)v_0^{-1/2}(\xi)]d\xi \quad \text{(A16)}$$

$$v_1(x) = v_0(x) + \int_0^x (\xi - x)[v_0''(\xi) - (\gamma/\beta)u_0^{1/2}(\xi)v_0^{1/2}(\xi) + \gamma^2/\beta^2\, u_0(\xi)]d\xi \quad \text{(A17)}$$

$$v_1(x) = v_0(x) + \int_0^x (\xi - x)[v_0''(\xi) - v_0(\xi) + (\beta/\gamma)v_0^{3/2}(\xi)u_0^{-1/2}(\xi)]d\xi \quad \text{(A18)}$$

Assuming that its initial approximate solutions which satisfies the boundary conditions (10), (11) have the form

$$u_0(x) = [1 + a(1 - x^2)]^2 \quad \text{(A19)}$$

$$v_0(x) = [1 - 2ax + ax^2]^2 \quad \text{(A20)}$$

By the iteration formula (A15) to (A18) we have the **Eqs.18**, **21**, **26** and **29** in the text.

## APPENDIX B

Scilab Program to find a solution of the **Eqs.47-51**.

```
function pdex4
m = 0;
x = linspace(0,1);
t=linspace(0,1);
sol = pdepe(m,@pdex4pde,@pdex4ic,@pdex4bc,x,t);
u1 = sol(:,:,1);
u2 = sol(:,:,2);
figure
surf(x,t,u1)
title('u1(x,t)')
xlabel('Distance x')
ylabel('Time t')
figure
surf(x,t,u2)
title('u2(x,t)')
xlabel('Distance x')
ylabel('Time t')
% --------------------------------------------------------------
function [c,f,s] = pdex4pde(x,t,u,DuDx)
c = [1; 1];
f = [1; 1] .* DuDx;
y = u(1) * u(2);
gamma=0.01;
beta=0.01;                                  % parameters
F  =(gamma*beta*sqrt(y))/(1+gamma*beta*(sqrt(u(2)/
u(1))));
F1=(gamma*sqrt(y))/( beta+gamma*(sqrt(u(1)/u(2))));
% non linear terms
s=[-F;-F1];
% --------------------------------------------------------------
function u0 = pdex4ic(x); %create a initial conditions
u0 = [1; 0.001];
% --------------------------------------------------------------
function [pl,ql,pr,qr] = pdex4bc(xl,ul,xr,ur,t) %create
a boundary conditions
pl = [0; ul(2)-1];
ql = [1; 0];
pr = [ur(1)-1; 0];
qr = [0; 1];
```

**Vol.2, No.6, 626-630 (2010)**
doi:10.4236/ns.2010.26077

**Natural Science**

# Thermophysical properties of dunite rocks as a function of temperature along with the prediction of effective thermal conductivity

**Aurang Zeb[1*], Tayyaba Firdous[2], Asghari Maqsood[3]**

[1]Department of Applied Physics, Federal Urdu University of Arts Science and Technology, Islamabad, Pakistan; *Corresponding Author: aurangzebbabar@yahoo.com
[2]Department of Physics, Air University, Islamabad, Pakistan;
[3]Thermal Transport Laboratory, School of Chemical and Materials Engineering, NUST, Islamabad, Pakistan; asgharimaqsood@yahoo.com

Received 29 December 2009; revised 4 March 2010; accepted 7 April 2010.

## ABSTRACT

**The thermal conductivity, thermal diffusivity and heat capacity per unit volume of dunite rocks taken from Chillas near Gilgit, Pakistan have been measured simultaneously using transient plane source technique. The temperature dependence of thermal transport properties is studied in the temperature range 83-303 K. Different relations for the estimation of thermal conductivity are also tested. Thermal conductivity data obey the modified Eucken's law in the temperature range of measurements.**

**Keywords:** Dunite; Density; Porosity; Thermal Conductivity; Transient Plane Source (TPS) Technique

## 1. INTRODUCTION

The most relevant thermal parameters of rocks are thermal conductivity, heat capacity per unit volume and thermal diffusivity. The first two parameters give the capability of a material to conduct and accumulate heat, respectively; and the last one represents how fast heat diffuses through a material [1].

### 1.1. Introduction to Samples

Igneous rocks are classified on the basis of texture and chemical composition.

On the basis of texture igneous rocks are classified in to two groups.

#### 1.1.1. Extrusive or Volcanic

When magma comes out of the surface of earth (called lava) then rapid cooling and crystallization of that magma above the earth surface form volcanic igneous rocks. Most types of lavas cool rapidly, resulting in the formation of rocks composed mainly of microscopic crystals. Some lavas cool so quickly that they form a smooth volcanic glass called obsidian. These are too fine-grained or glassy that their mineral composition cannot be observed without the use of petrographic microscope.

#### 1.1.2. Intrusive or Plutonic

The cooling and crystallization of molten magma below the surface of earth form these rocks. Magma that forms the intrusive rocks solidifies relatively slow; and so, most of the intrusive rocks have larger crystals than of extrusive rocks. The mineral grain size of these rocks is visible to naked eye even.

On the basis of chemistry igneous rocks are classified in to four groups [2].

#### 1.1.3. Felsic

Igneous rocks derived from felsic magma contain relatively high quantities of sodium, aluminium and potassium and are composed of more than 65% silica ($SiO_2$). Some common felsic igneous rocks include fine-grained Rhyolite and coarse-grained Granite. All of felsic rocks are light in colour because of the dominance of quartz, potassium and sodium feldspars, and plagioclase feldspars minerals.

#### 1.1.4. Intermediate

Some igneous rocks having chemistry between felsic and mafic rocks are known as intermediate. Silica amounts from 52% to 65%. Andesite (intrusive) and Diorite (extrusive) are intermediate igneous rocks. These rocks are composed predominantly Plagioclase feldspar, Amphibole and Pyroxene minerals.

#### 1.1.5. Mafic

Igneous rocks derived from mafic magma rich in cal-

cium, iron and magnesium and relatively poor in silica, amounts from 45% to 52%. Some common mafic igneous rocks include-fine grained Basalt and coarse-grained Gabbro. Mafic igneous rocks tend to be dark in colour because they contain a large proportion of minerals rich in iron and magnesium (Pyroxene, Amphiboles and Olivine).

### 1.1.6. Ultramafic

These rocks contain relatively low amount of silica < 45% and are dominated by the minerals olivine, calcium rich plagioclase feldspars and pyroxene. Peridotite and Dunite are the most common ultramafic rocks having coarse-grained texture. There is no known modern fine-grained ultramafic rock. The samples studied here belong to the dunite group of igneous rocks.

The density related properties of the selected samples at room temperature along with their thermal properties in temperature range (303-483 K) are already published [3]. In this paper only the thermal transport properties are measured in temperature range 83 to 303 K at normal pressure, with air as saturant in pore spaces, using Transient Plane Source Technique [4].

In continuation of our previous work [5,6], the aim in this work is to study thermal transport properties of dunite rocks as a function of temperature and to test various relations for the prediction of effective thermal conductivity of porous media.

## 2. EXPERIMENTAL TECHNIQUES AND SAMPLE CHARACTERIZATION

The thermal transport properties of the samples were measured as a function of temperature, using transient plane source (TPS) technique. The beauty of TPS technique is that it allows measurements without any disturbance from the interfaces between the sensor and the bulk samples. Also, simultaneous measurement of thermal conductivity and thermal diffusivity is possible [4]. In this technique, a TPS-element (**Figure 1**) sandwiched between two halves of the sample is used both as a constant heat source and a sensor of temperature.

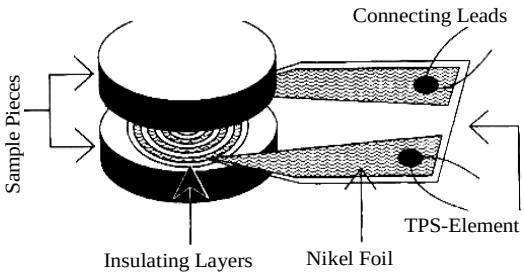



**Figure 1.** A TPS-sensor sandwiched between sample pieces [7].

For data collection the TPS-element (20 mm diameter) was used in a bridge circuit, shown in **Figure 2**.

When sufficiently large (constant) amount of direct current is passed through the TPS-element, its temperature changes consequently and there is a voltage drop across the TPS-element. By recording this voltage drop for a particular time interval, detailed information about the thermal conductivity ( $\lambda$ ) and thermal diffusivity ( $\kappa$ ) of the test specimen is obtained. The heat capacity per unit volume ( $\rho C_P$ ) can then be calculated from the relation:

$$\rho C_P = \frac{\lambda}{\kappa}, \tag{1}$$

The results of the thermophysical measurements on the samples at different temperatures and normal pressure are shown in **Figure 3**.

It is to be noted that the results of thermal properties in high temperature range 303 to 483 K [3] are also included in **Figure 3**, so that one can observe the overall behavior of thermal properties of these samples over a wide range of temperature (83-483 K).

Taking into consideration the errors of the technique [7,8], standard deviations of the measurements and the sampling errors, the thermal conductivity and thermal diffusivity data contain errors of 5% and 7% respectively. The errors in volumetric heat capacity are around 10%.

## 3. THERMAL CONDUCTIVITY PREDICTION

A typical equation used for temperature dependence of lattice (phonon) thermal conductivity is:

$$\frac{1}{\lambda(T)} = A + BT\ , \tag{2}$$

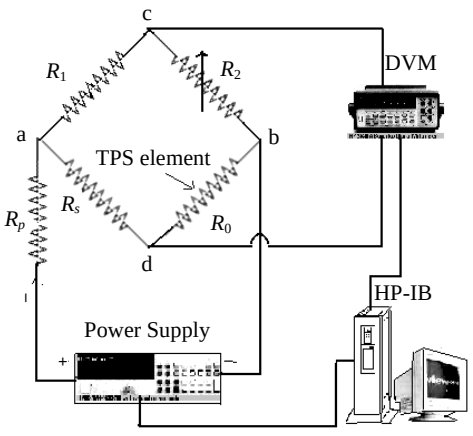



**Figure 2.** Bridge circuit diagram for TPS technique: $R_s$ = Standard resistance, $R_0$ = Resistance of TPS sensor [9].

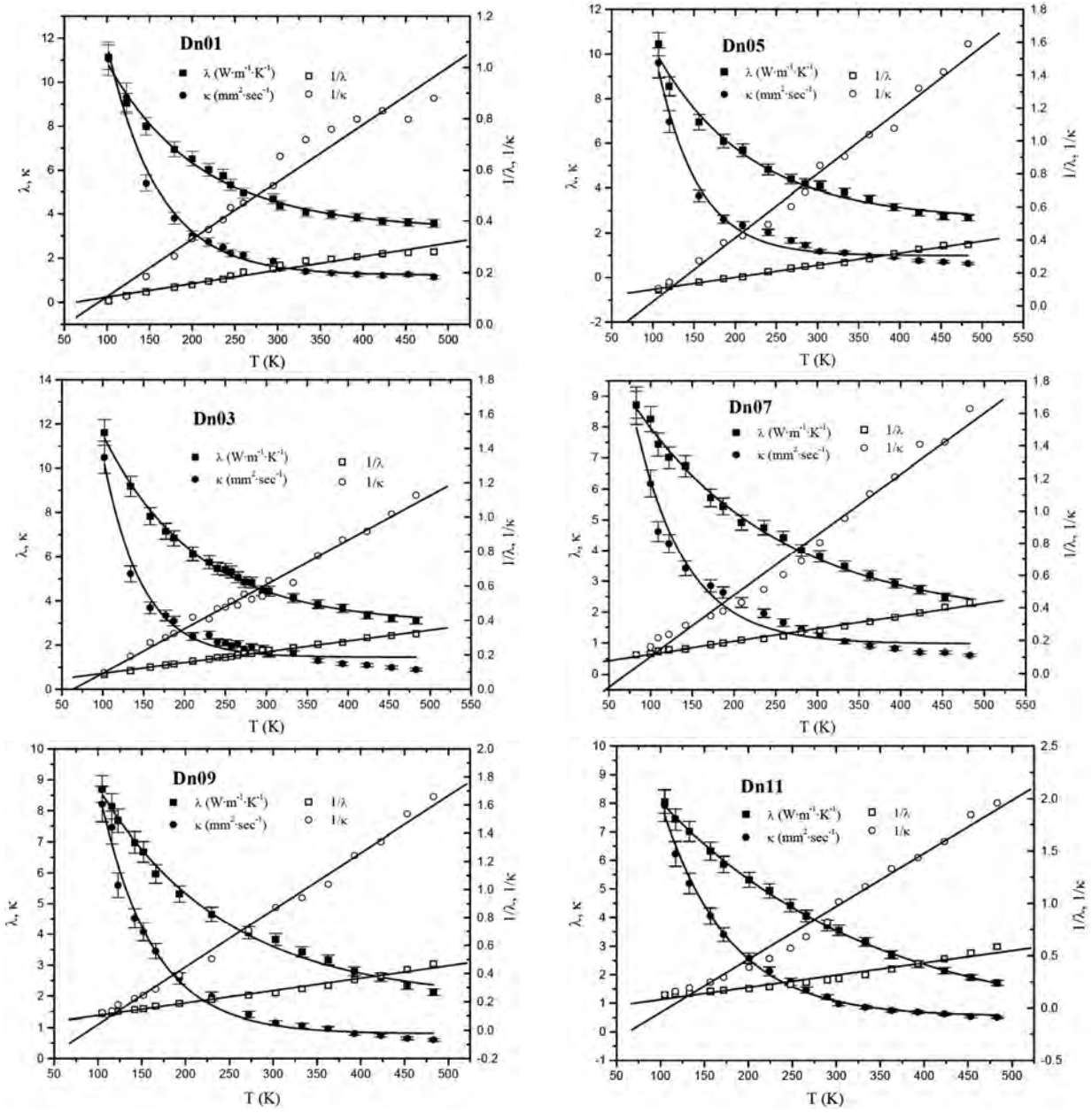



**Figure 3.** Temperature dependence of thermal conductivity, thermal diffusivity and heat capacity per unit volume, $\frac{1}{\lambda(T)} = A + BT$ and $\frac{1}{\kappa(T)} = R + ST$. Estimated uncertainties in $\lambda$ and $\kappa$ are about 5% and 7% respectively.

where $A$(W$^{-1}$·m·K) and $B$(W$^{-1}$·m) are constants related to the scattering properties of phonons. $A$ is related to scattering of phonons by impurities and imperfections [10]; and $B$ is related to phonon-phonon scattering and is approximately proportional to an inverse power of sound velocity [11].

The physical justification of the term $A$ is the existence of numerous additional scattering centers for phonons in materials, caused by structural and chemical imperfections and the influence of the grain boundaries.

**Eq.2** is analogous to Matthiessen's rule of electrical resistivity in metals, which is:

$$\rho_e = \rho_{eo} + \rho_e(T), \quad (3)$$

where $\rho_{eo}$ is the extrapolated electrical resistivity at 0 $K$ and is called the residual resistivity. This is the temperature independent part of the resistivity. That is, the value of $\rho_{eo}$ depends upon the concentration of impurities and other imperfections in the sample. It can be taken as the measure of impurity of the specimen. The tempera-

ture dependent part of resistivity $\rho_e(T)$ is decided by the phonons. It is called the lattice resistivity [12].

For diffusivity (having the same trend as thermal conductivity), the analogous equation was assumed to be:

$$\frac{1}{\kappa(T)} = R + ST \tag{4}$$

The results obtained using **Eqs.2** and **4** along with the experimental results / values of $\lambda$ and $\kappa$ are shown in **Figure 3**. The correlation coefficients for thermal conductivity $(r_\lambda)$ and thermal diffusivity $(r_\kappa)$ are also listed in **Table 1**.

The data appear to fit very well (above 97%). The values of $\rho C_p$ can be calculated using **Eq.1**.

**Eq.2** may also be written as [13,14]:

$$\lambda(T) = \lambda_r(1 + bT_r)/(1 + bT), \tag{5}$$

where $\lambda_r$ ($W \cdot m^{-1} \cdot K^{-1}$) is the thermal conductivity at a reference temperature and $b(K^{-1})$ is a single temperature coefficient of thermal conductivity parameter, controlling the temperature dependence of thermal conductivity. $b$ is related to parameters $A$ and $B$ of **Eq.2**, as:

$$b = B/A \tag{6}$$

The reference value of thermal conductivity ($\lambda_r$) is simply derived from **Eq.2**, using $T = T_r$ (reference/ room temperature). It is to be noted that each of the **Eq.2** or 5 is the modified Eucken's rule.

A recently proposed empirical model [6] for the prediction of thermal conductivity as a function of temperature is:

$$\frac{1}{\lambda_e} = \frac{1}{\lambda_s} + \frac{m\Phi}{\lambda_f}\left(\frac{T}{T_r}\right), \tag{7}$$

where $m$ is the empirical coefficient whose value can be determined at suitable temperatures, using the corresponding experimental values of thermal conductivity and the room temperature values of $\Phi$, $\lambda_f$ and $\lambda_s$; by

$$m = \lambda_f T_r \left( \frac{\sum\left(1/\lambda_{\exp} - 1/\lambda_s\right)}{\sum\left(T\Phi\right)} \right) \tag{8}$$

The empirical coefficients, exponents or adjustable parameters may vary according to the suite of rocks. Therefore, the extrapolations of empirical models to suites of rocks other than those used in developing these models may not be satisfactory [15].

## 4. RESULTS AND DISCUSSION

It is well known that thermal transport properties of porous rocks depend upon their structure, mineral composition, porosity, density, the ability of their constituent minerals to conduct heat, temperature, pressure; etc.

The temperature dependence of thermal conductivity and thermal diffusivity in the temperature range 83 to 483 K at suitable intervals is shown in **Figure 3**. It is observed that thermal conductivity decreases in the measured temperature range. This is in agreement with the theory of thermal conductivity. The thermal diffusivity shows decreasing trend; again in agreement with the theory.

For dunite samples, in **Eq.5**, the empirical coefficient b is taken to be 0.0172 which is equal to the mean of all the values of $b$ listed in **Table 1**. Similarly the mean value of $\lambda_r$ is found to be 3.858 $W \cdot m^{-1} \cdot K^{-1}$ at 303 K. Using these values of adjustable parameters $\lambda_r$ and $b$, thermal conductivity values are predicted along with the experimental measurements and are plotted in **Figure 4**.

It is to be noted that in **Eq.5**, all the parameters effecting thermal conductivity (such as porosity, density, thermal conductivity values of constituent minerals and the fluid inside the pores, etc.) are dumped into adjustable parameters and the errors in predicting thermal conductivity are only up to 10% in the temperature range 243 K to 333 K, which is the most interesting range for construction purposes.

**Table 1.** Parameters $A$, $B$, $R$ and $S$ for the thermal resistivity $1/\lambda(T)$ and the inverse of thermal diffusivity $1/\kappa(T)$ of dunites as represented by **Eqs.2** and **4**. $r_\lambda$ and $r_\kappa$ are the respective correlation coefficients.

| Specimen | $A$ $W^{-1} \cdot m \cdot K$ | $B$ $10^{-4} W^{-1} \cdot m$ | $r_\lambda$ | $R$ $m^{-2} \cdot sec$ | $S$ $10^{-3} m^{-2} \cdot sec \cdot K^{-1}$ | $r_\kappa$ | $b = B/A$ $K^{-1}$ | $\lambda_r$ $W \cdot m^{-1} \cdot K^{-1}$ |
|---|---|---|---|---|---|---|---|---|
| Dn01 | 0.053 | 5.24 | 0.98 | –0.116 | 2.23 | 0.98 | 0.010 | 4.728 |
| Dn03 | 0.029 | 6.29 | 1.00 | –0.165 | 2.58 | 0.99 | 0.021 | 4.545 |
| Dn05 | 0.025 | 7.36 | 1.00 | –0.362 | 3.88 | 0.99 | 0.029 | 4.027 |
| Dn07 | 0.040 | 7.75 | 1.00 | –0.271 | 3.74 | 0.99 | 0.019 | 3.639 |
| Dn09 | 0.016 | 8.80 | 0.99 | –0.365 | 4.06 | 1.00 | 0.056 | 3.545 |
| Dn11 | –0.036 | 1.17 | 0.98 | –0.550 | 5.06 | 0.99 | –0.032 | 3.140 |

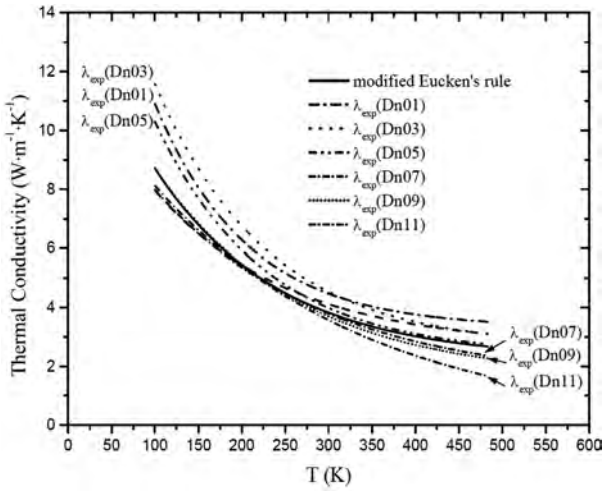



**Figure 4.** The comparison of experimental and predicted thermal conductivity of dunite samples as a function of temperature.

As far as the results of **Eq.7** are concerned, these are fairly good at room temperature and above (with in 10%) [3]. But, this proposal is not recommended for the prediction of thermal conductivity in temperature range lower than the room temperature. This is the limitation of this model. It is because of the fact that we have used the room temperature values of $\lambda_s$, $\lambda_f$ and $\Phi$ in **Eqs.7** and **8**. This makes the value of $m$ (calculated by **Eq.8**) negative in the temperature range lower than room temperature. This reverses the effect of second term on R.H.S. of **Eq.7**. As a consequence, we get the increasing effect in thermal conductivity with temperature due to this term which is against experimental results.

## 5. CONCLUSIONS

To predict the thermal conductivity of dunite samples as a function of temperature, modified Eucken's rule has been used. It is noted that the experimental thermal conductivity and the predicted conductivity are in good agreement (up to 10%) in the temperature range 243 K to 333 K, which is important for construction purposes.

## 6. ACKNOWLEDGEMENTS

The authors wish to thank Mr. Zulqurnain Ali, Mr. Matloob Hussain (Earth Science Department, QAU) and Mr. Muhammad Saleem Mughal (Statistics Department, QAU) for helpful discussion. The author (Aurang Zeb) is grateful to Higher Education Commission, Pakistan for supporting the doctoral studies financially.

## REFERENCES

[1] Cengel, Y.A. and Turner, R.H. (2001) Fundamentals of thermal-fluid sciences. McGraw-Hill, Boston, 609.

[2] Rzhevsky, V. and Novik, G. (1971) The physics of rocks. Mir Publishers, Moscow.

[3] Aurangzeb, Mehmood, S. and Maqsood, A. (2008) Modeling of effective thermal conductivity of dunite rocks as a function of temperature. *International Journal of Thermophysics*, **29(4)**, 1470-1479.

[4] Gustafsson, S.E. (1991) Transient plane source techniques for thermal conductivity and thermal diffusivity measurements of solid materials. *Review of Scientific Instruments*, **62(3)**, 797-804.

[5] Aurangzeb, Ali, Z., Gurmani, S.F. and Maqsood, A. (2006) Simultaneous measurement of thermal conductivity, thermal diffusivity and prediction of effective thermal conductivity of porous consolidated igneous rocks at room temperature. *Journal of Physics D*: *Applied Physics*, **39(17)**, 3876-3881.

[6] Aurangzeb, Khan, L.A. and Maqsood, A. (2007) Prediction of effective thermal conductivity of porous consolidated media as a function of temperature: A test case of limestones. *Journal of Physics D*: *Applied Physics*, **40(16)**, 4953-4958.

[7] Maqsood, M., Arshad, M., Zafarullah, M. and Maqsood, A. (1996) Low-temperature thermal conductivity measurement apparatus: Design assembly, caliberation and measurement on (Y123, Bi2223) superconductors. *Superconductor Science and Technology*, **9(4)**, 321-326.

[8] Maqsood, A. and Rehman, M.A. (2003) Measurement of thermal transport properties with an improved transient plane source technique. *International Journal of Thermophysics*, **24(3)**, 867-883.

[9] Maqsood, A., Rehman, M.A., Gumen, V. and Haq, A. (2000) Thermal conductivity of ceramic fibres as a function of temperature and press load. *Journal of Physics D*: *Applied Physics*, **33(16)**, 2057-2063.

[10] Schatz, J.F. and Simmons, G. (1972) Thermal conductivity of earth materials at high temperatures. *Journal of Geophysical Research*, **77(35)**, 6966-6983.

[11] Ziman, J.M. (1960) Electrons and phonons. Oxford University Press, London.

[12] Vijaya, M.S. and Rangarajan, G. (2004) Materials science. McGraw-Hill, New Delhi.

[13] Kukkonen, I.T., Jokinen, J. and Seipold, U. (1999) Temperature and pressure dependencies of thermal transport properties of rocks: Implications for uncertainties in lithosphere models and new laboratory measurements of high-grade rocks in the Central Fennoscandian Shield. *Surveys in Geophysics*, **20(1)**, 33-59.

[14] Jokinen, J. (2000) Uncertainty analysis and inversion of geothermal conductive models using random simulation methods. Oulu University, Oulu.

[15] Somerton, W.H. (1992) Thermal properties and temperature related behaviour of rock/fluid systems. Elsevier, New York.

**Vol.2, No.6, 631-634 (2010)**
doi:10.4236/ns.2010.26078

**Natural Science**

# Optical properties for N,N'-bis (Inaphyhly)-N,N'-diphenyl-1,1'-biphenyl-4,4'-diamine and tris (8-hydroxyquinolinato) aluminum in organic light emitting devices

**Mei Yee Lim[1*], Wan Mahmood Mat Yunus[1], Zainal Abidin Talib[1], Anuar Kassim[2]**

[1]Department of Physics, Faculty of Science, University Putra Malaysia, Selangor, Malaysia; *Corresponding Author: mahmood@science.upm.edu.my
[2]Department of Chemistry, Faculty of Science, University Putra Malaysia, Selangor, Malaysia

Received 12 January 2010; revised 9 March 2010; accepted 2 April 2010.

## ABSTRACT

**The optical properties of N,N'-bis (Inaphthyl)-N,N'-diphenyl-1,1'-biphenyl-4,4'-diamine (NPB) and tris (8-hydroxyquinolinato) aluminum ($Alq_3$) organic materials used as hole transport and electron transport layers in organic light-emitting devices (OLED) have been investigated. The NPB and $Alq_3$ layers were prepared using thermal evaporation method. The results show that the energy band gap of $Alq_3$ is thickness independence while the energy band gap of NPB decreases with the increasing of sample thickness. For the case of photoluminescence the $Alq_3$ with thickness of 84 nm shows the highest relative intensity peak at 510 nm.**

**Keywords:** Energy Band Gap; N,N'-Bis (Inaphthyl) -N,N'-Diphenyl-1,1'-Biphenyl-4,4'-Diamine Tris (8 Hydroxyquinolinato) Aluminum; Organic Light Emitting Diode; Photoluminescence

## 1. INTRODUCTION

Various organic light emitting devices displays have been investigated in recent years. Many techniques like anode [1] and cathode [2] modifications, inserting hole blocking layer [3], control the thermal evaporation rate [4] and doping system have been recently reported to further improve display efficiency due to their excellent performances. Currently, there is great interest in the study of OLEDs containing small molecules as emitting layer. Such molecules have been found to be useful in OLED applications such as optical devices, luminescence probes in biomedical assays, luminescence sensors for chemical species and fluorescent lighting [5,6]. Advantages of organic LEDs over inorganic ones are their high emission efficiency in the visible spectrum, easy to process, robustness and almost infinite possibility for modification. They also have low operating voltage, low power consumption and wide viewing angle. However, there are still several problems which must be solved especially for transferring research results to commercial application of OLED devices in display technology. Short OLED lifetime, low carrier mobility and high interface barriers must be improved by optimizing the material parameters and fabrication steps.

As present, the basic structure of an OLED is shown in **Figure 1** and typically consists of three organic semiconductor layers (hole injection layer, organic emitter and electron transport layer) sandwiched between two electrodes. The electron-injecting electrode consists of a

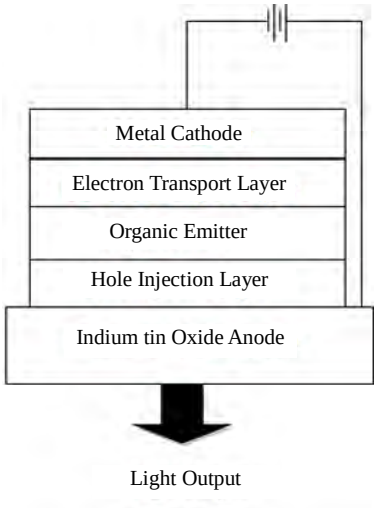



**Figure 1**. The basic structure for organic light emitting device.

low work function metal alloy, typically Mg-Ag or Li-Al, deposited by vacuum evaporation. The bottom, hole-injecting, electrode is typically a thin-film of the transparent semiconductor indium tin oxide (ITO) [7]. Upon recombination, energy is released as light, which is emitted from the light-transmissive anode and substrate. N'-bis (Inaphthyl)-N,N'-diphenyl-1,1'-biphenyl-4,4'-diamine (NPB) was chosen as hole injection layer because it can be easily manufactured and is abundantly available in powder form. However, tris (8-hydroxyquinolinato) aluminum ($Alq_3$) is used as an ETL because $Alq_3$ is thermally and morphologically stable therefore can be easily evaporated into thin films form. Easily synthesized and purified, molecularly shaped to avoid exciplex formation (e.g. with N,N'-bis (Inaphthyl)-N,N'-diphenyl-1,1'-biphenyl-4,4'-diamine at the interface), and produce green fluorescent light become another reason to be a good host emitter [8].

## 2. EXPERIMENTAL

For the preparation of $Alq_3$ and NPB organic layers, the indium tin oxide (ITO) glasses substrate was cut into square plates (2 cm × 2 cm). The ITO glasses were immersed in ultrasonic baths with acetone for 10 minutes. Then, the ITO glasses were rinsing in deionised water for 10 min and then blow dried with nitrogen gas. This procedure was applied to remove organic contamination and particles from the ITO surface. Thermal evaporation technique using resistively heated tantalum boats in vacuum, at a base pressure of $1.0 \times 10^{-5}$ Torr was used to prepare thin film sample. Different thickness of organic layers was deposited on the ITO at a rate of 2.5 A/s at room temperature. The thicknesses of the layers were measured by Tencor P-12 Disk Profile. The PL spectra of the devices were measured by a EL spectra USB 2000 FLG Spectrofluorometer. The spectra of optical absorption measurements were made over the wavelength range of 360 nm-800 nm to obtain the energy band gap.

## 3. RESULTS AND DISCUSSION

### 3.1. Energy Band Gap

The spectra of optical absorption measurements were made over the wavelength range of 360 nm-800 nm. The variation in the absorbance with wavelength is shown in **Figures 2** and **3** for $Alq_3$ and NPB with varying thickness.

The optical absorption spectra of both samples, $Alq_3$ and NPB are similar thus the optical band gap was calculated using a well known equation as

$$(\alpha h\nu)^2 = c\left(h\nu - E_g\right) \quad (1)$$

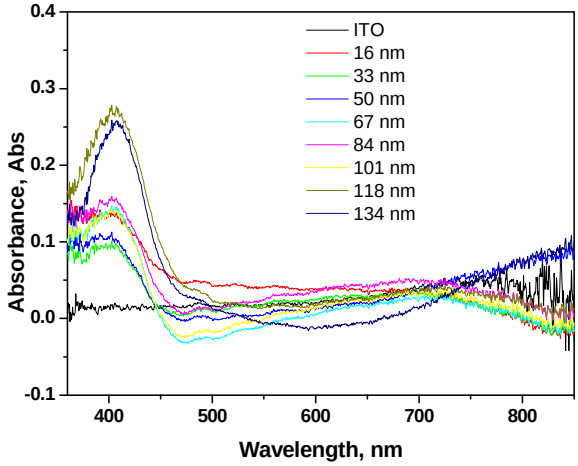



**Figure 2**. The optical absorption of $Alq_3$ at various thickness.

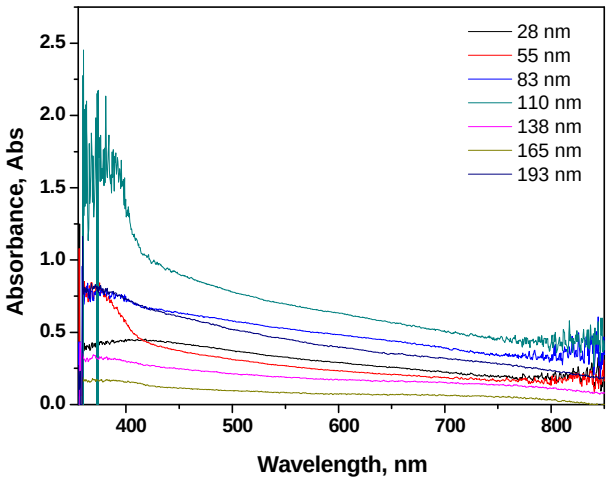



**Figure 3.** The optical absorption of NPB measured at different thicknesses.

where $c$ is a constant, $E_g$ is the optical band gap, $h\nu$ is photon energy. The energy band gap was obtained by plotting $(\alpha h\nu)^2$ as a function of photon energy, $h\nu$.

The result shows that the energy band gap for the $Alq_3$, is tend to be independent on the layer thickness as shown in **Figure 4**. However for NPB layers the energy band gap decreases with the increasing NPB thickness as displayed in **Figure 5**. This result can be easily correlated with the efficient hole transport in hole injection layer as discussed by Zhang Zhi-Feng *et al*. This phenomenon is due to a good balance between the injected electrons and holes in the OLED structure [9].

### 3.2. Photoluminescence

The photoluminescence spectra of $Alq_3$ were successfully measured using Ocean Optics spectrofluorometer operated at 390 nm. **Figure 6** shows the normalized photoluminescence intensity for samples at various thick-

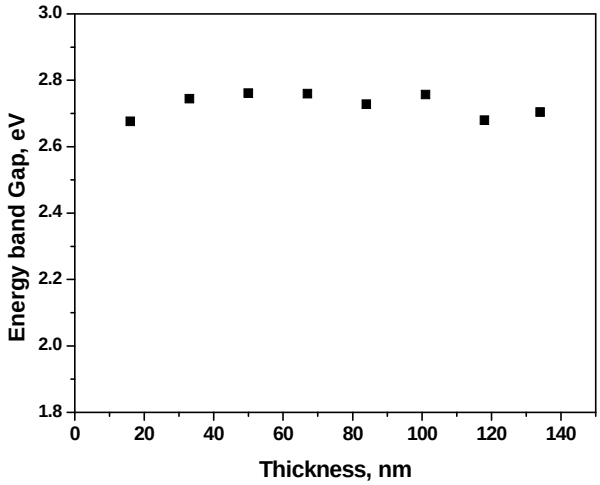



**Figure 4.** Energy band gap for different thickness of $Alq_3$.

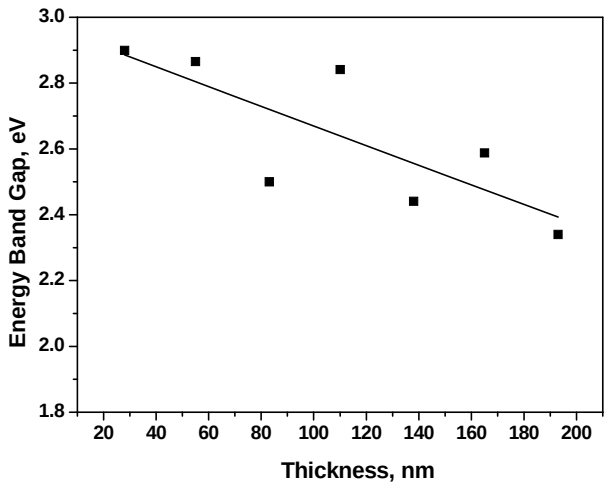



**Figure 5.** Energy band gap for different thickness of NPB.

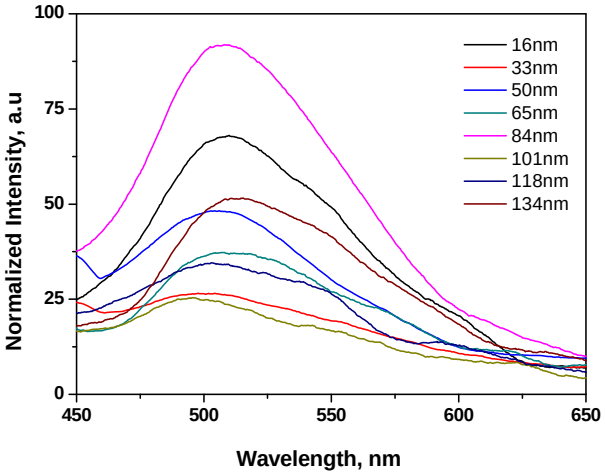



**Figure 6.** The PL performance with the different thickness of $Alq_3$.

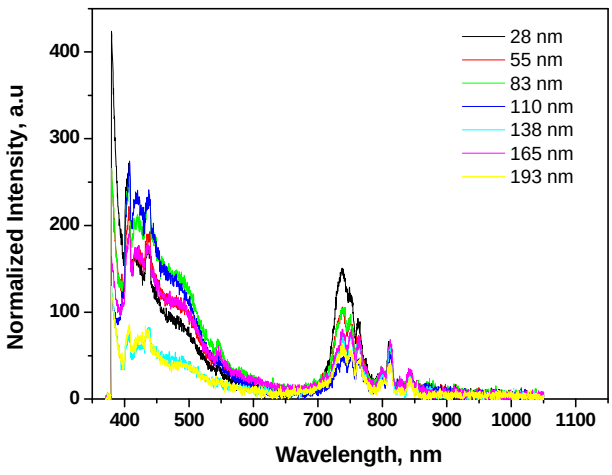



**Figure 7.** The PL performance with the different thickness of NPB.

nesses. It can be observed that $Alq_3$ with 84 nm thickness gives the highest relative intensity at emission wavelength of 510 nm. In the case of NPB layer there is no any photoluminescence phenomenon occurred as displayed by **Figure 7**. The spectra in **Figure 7** are the optical reflectance of the excitation source. However NPB layer with a thickness 110 nm gives the highest relative reflectance intensity at the peak of excitation source, *i.e.* 390 nm and 750 nm respectively.

## 4. CONCLUSIONS

The energy band gap of $Alq_3$ and NPB layers for OLED structure have successfully been measured. The NPB layer shows a thickness dependence of energy band gap while the energy band gap for $Alq_3$ is tend to be independence on the sample thickness ranging from 16 nm to 134 nm. As electron transport layer in OLED structure, the photoluminescence spectrum of $Alq_3$ was observed at 510 nm with the optimum thickness of 84 nm.

## REFERENCES

[1] Van Slyke, S.A., Chen, C.H. and Tang, C.W. (1996) Organic electroluminescent devices with improved stability. *Applied Physics Letters*, **69(15)**, 2160-2162.

[2] Hung, L.S., Tang, C.W. and Mason, M.G. (1997) Enhanced electron injection in organic electroluminescence devices using an Al/LiF electrode. *Applied Physics Letters*, **70(2)**, 152-155.

[3] O'Brien, D.F., Baldo, M.A., Thompson, M.E. and Forrest, S.R. (1999) Improved energy transfer in electrophosphorescent devices. *Applied Physics Letters*, **74(3)**, 442-444.

[4] Lee, C.B., Uddina, A., Hua, X. and Andersson, T.G. (2004) Study of $Alq_3$ thermal evaporation rate effects on the

OLED. *Materials Science & Engineering B*, **112(1)**, 14-18.

[5] Kido, J. and Okamoto, Y. (2002) Organo lanthanide metal complexes for electroluminescent materials. *Chemical Reviews*, **102(6)**, 2357-2368.

[6] Silva, V.M. and Pereira, L. (2006) The nature of the electrical conduction and light emitting efficiency in organic semiconductors layers: The case of [m-MTDATA] – [NPB] – $Alq_3$ OLED. *Journal of Non-Crystalline Solids*, **352(50-51)**, 5429-5436.

[7] Burrows, P.E., Forrest, S.R. and Thompson, M.E. (1997) Prospects and applications for organic light-emitting devices. *Current Opinion in Solid State & Materials Science*, **2(2)**, 236-243.

[8] Hung, L.S. and Chen, C.H. (2002) Recent progress of molecular organic electroluminescent materials and devices. *Materials Science & Engineering R*, **39(5-6)**, 143-222.

[9] Zhang, Z.F., Deng, Z.B., Liang, C.J., Zhang, M.X. and Xu, D.H. (2003) Organic light-emitting diodes with a nanostructured $TiO_2$ layer at the interface between ITO and NPB layers. *Displays*, **24(4-5)**, 231-234.

**Vol.2, No.6, 635-640 (2010)**
doi:10.4236/ns.2010.26079

**Natural Science**

# The dynamic field in turbulent round jet discharging into a co-flowing stream

**Mohamed Hichem Gazzah[1*], Nejmiddin Boughattas[2], Hafedh Belmabrouk[1], Rachid Said[2]**

[1]Département de Physique, Faculté des Sciences de Monastir, Monastir, Tunisie; *Corresponding Author: Hichem.Gazzah@fsm.rnu.tn
[2]UREMIR, Institut Préparatoire aux Etudes d'Ingénieurs de Monastir, Monastir, Tunisie

Received 22 February 2010; revised 16 March 2010; accepted 7 April 2010.

## ABSTRACT

**The effects of a co-flow on a spreading and entrainment rate of turbulent round jets have been studied numerically. The first and second order closure models are used and have been compared with existing experimental data. The influence of theses models on the dynamic fields is examined. The results of the models in general agree well with the trends observed experimentally. The co-flowing imposed noticeable restrictions on the spreading and the turbulent mixing. Finally, an entrainment hypothesis has been introduced to describe the development of turbulent jets issuing into a stagnant or co-flowing air. It relates the mass flow rate of the surrounding fluid entrained into the jet to the characteristic velocity difference between the jet and the co-flow. It is obvious that the co-flow decreases considerably the entrainment of air.**

**Keywords:** Co-Flow; Turbulence; Jets; Models; Entrainment

## 1. INTRODUCTION

The experimental and numerical studies concerning the aerodynamics of co-flowing turbulent round jets have lately observed an increasing interest as they are encountered in several industrial applications, such as turbulent diffusion flames of combustion chambers where a fuel jet flow is commonly injected into a co-flowing stream jet. The important parameters that influence the mixing characteristics of a jet are the presence of density difference and a co-flowing between the jet and its surroundings.

The effect of the density variation in the turbulent round free jet has been investigated experimentally and numerically by several authors like [1-6]. The results of these works show that the mean and turbulent quantities are strong functions of density ratio. This confirms a higher mixing efficiency when the density ratio between the jet and the quiescent air decreases. It is shown that the density effects are affected by the buoyancy terms in the similarity region of the jet. The influence of the emission initial conditions on the evolution of the dynamic and scalar fields is also studied.

The present study is a continuation of the work started by the authors [7,8] to investigate numerically the influence of the co-flow surrounding the turbulent jets. In this numerical study, several traditional scalar dissipation rate models are examined for scalar transport modeling in mixing turbulent round jets with co-flowing air. Therefore, it is not intended here to review previous work detail [9-11] on a co-flowing jet, since a review on the subject matter is given by the references [7,8]. The experimental and numerical investigations of this type of flow are also relatively scarce.

In his recent work, Hakem *et al.* [12], have studied experimentally the mixing characteristics of an elliptical jet with large varying aspect ratio in a co-flow current, to verify that the Elliptic jet with varying large aspect ratio has also much higher dilution in a co-flow than an equivalent round jet under the same flow conditions. Wang *et al.* [13] have also studied the variable-density turbulent round jets discharging into a weakly confined low-speed co-flowing air stream with the aid of large-eddy simulation. Thus, the majority of this work shows that the effects of the co-flow variation on the structure of the reactive and non-reactive turbulent jets are complex problems and remain very interesting.

The study of the jets with co-flow shows the presence of three essential areas: an initial area, a principal area and an area of transition. When the flow is confined, the process of the co-flow driven by the jet is modified and the mixing process depends strongly not only on the velocity ratio, but also on the interaction between the boundary layer, the mixing layer and the main flow.

A considerable pressure gradient can appear and generates phenomena of recirculation. Curtet [14] is interested in a parameter of similarity, called parameter of Craya-Curtet which is formulated in the literature in variable density by Steward and Gurus [15]. He showed that, for a value of this number higher than 0.8, the phenomenon of recirculation is avoided, irrespective of the fluid considered.

The major objective of this paper is to determine the effect of a co-flowing on the dynamic fields of a turbulent round jet. Therefore, we investigate a turbulent round jet into a co-flowing air with various co-flow to fuel velocity ratios. We have thus used first and second order closure methods to investigate and compare their performances. The influence of the co-flow on various physical parameters of the jet is analyzed in comparison with the experimental data of Djeridane [16].

## 2. TURBULENCE MODELS

The equations which govern the turbulent flow are derived from the conservation laws of mass and momentum. All variables are conventionally averaged. These conventional averaged variables are denoted by an overbar $\overline{\Phi}$. Conventional fluctuations are indicated by $\Phi'$.

$$\Phi = \overline{\Phi} + \Phi' \tag{1}$$

### 2.1. The Mean Equations

We suppose that, the mean motion is steady, the turbulent Reynolds numbers are high enough and the molecular diffusion effects are neglected.

The continuity equation is given by

$$\rho \frac{\partial \overline{U}_j}{\partial x_j} = 0 \tag{2}$$

The momentum equation is:

$$\rho \frac{\partial}{\partial x_j}\left(\overline{U}_i \overline{U}_j\right) = -\frac{\partial \overline{p}}{\partial x_i} - \rho \frac{\partial}{\partial x_j}\left(\overline{u_i' u_j'}\right) + \rho g_i \tag{3}$$

As a consequence of the nonlinearity (2), the averaging process used introduces unknown correlations which are modelled through turbulence models. In order to solve transport equation for mean velocity in the turbulent jets, the turbulent Reynolds stress shown in these equations is computed using two turbulence closure models, called the *k*-*ε* model and the second order model. Details of theses models can be found in Schiestel [17].

### 2.2. The First-Order *k*-*ε* Model

The turbulent fluxes are approximated with *k*-*ε* model. The Reynolds stresses tensors are related to the strain rate by the following equation:

$$\overline{\rho u_i' u_j'} = -\mu_t \left( \frac{\partial \overline{U}_i}{\partial x_j} + \frac{\partial \overline{U}_j}{\partial x_i} \right) + \frac{2}{3} \overline{\rho} k \delta_{ij} \tag{4}$$

where $\mu_t$ is the turbulent viscosity, which is obtained from the turbulent kinetic energy $k$ and its dissipation rate $\varepsilon$ using the relation:

$$\mu_t = \rho C_\mu \frac{k^2}{\varepsilon} \tag{5}$$

The turbulence model consists of equations for the turbulent kinetic energy and their dissipation. These are

The kinetic energy equation

$$\rho \frac{\partial}{\partial x_j}\left(\overline{U}_j\, k\right) = \frac{\partial}{\partial x_j}\left( \frac{\mu_t}{\sigma_k} \frac{\partial k}{\partial x_j} \right) - \rho \overline{u_i' u_j'} \frac{\partial \overline{U}_i}{\partial x_j} - \rho \varepsilon \tag{6}$$

The energy dissipation rate

$$\rho \frac{\partial}{\partial x_j}\left(\overline{U}_j\, \varepsilon\right) = \frac{\partial}{\partial x_j}\left( \frac{\mu_t}{\sigma_\varepsilon} \frac{\partial \varepsilon}{\partial x_j} \right) + Ce_1 P_k \frac{\varepsilon}{k} - Ce_2 \rho \frac{\varepsilon^2}{k} \tag{7}$$

The model constants used in the present study are given in **Table 1**.

The *k*-*ε* model has been used with success in the calculation of various turbulent jets. However, in flows with significant streamline curvature, the isotropic eddy viscosity assumption may not be able to describe the turbulent diffusion effects adequately.

### 2.3. The Second-Order Model

The second turbulence model considered in this study is a Reynolds Stress Model (RSM). The Reynolds stress equation is:

$$\rho \frac{\partial}{\partial x_k}\left(\overline{U}_k \overline{u_i' u_j'}\right) = P_{ij} + D_{ij} + \overline{p'\left( \frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i} \right)} - \frac{2}{3} \rho \varepsilon \delta_{ij} \tag{8}$$

The first term on the right hand side is the production term due to the mean strain:

$$P_{ij} = -\rho \overline{u_i' u_k'} \frac{\partial \overline{U}_j}{\partial x_k} - \rho \overline{u_j' u'}_k \frac{\partial \overline{U}_i}{\partial x_k} \tag{9}$$

The diffusion term is modelled as:

$$D_{ij} = C_s \rho \frac{\partial}{\partial x_k}\left( \rho \frac{k}{\varepsilon} \overline{u_k' u_l'} \frac{\partial \left(\overline{u_j' u_j'}\right)}{\partial x_l} \right) \tag{10}$$

**Table 1.** Turbulence constants for the first order *k*-*ε* model, where the value of $C_\mu$ is adapted for the axisymmetric jet case.

| $C\mu$ | $C_{e,1}$ | $C_{e,2}$ | $\sigma_k$ | $\sigma_\varepsilon$ |
|---|---|---|---|---|
| 0.06 | 1.44 | 1.92 | 1.00 | 1.30 |

The pressure-strain correlation is:

$$\overline{p'\left(\frac{\partial u_i'}{\partial x_j}+\frac{\partial u_j'}{\partial x_i}\right)}=\phi_{ij}^{I}+\phi_{ij}^{II} \quad (11)$$

$\phi_{ij}^{I}=-C_1\rho\varepsilon\left(\rho\frac{\overline{u_j'u_j'}}{k}-\frac{2}{3}\delta_{ij}\right)$: The return to isotropy term.

$\phi_{ij}^{II}=-C_2\left(P_{ij}-\frac{2}{3}P_{kk}\delta_{ij}\right)$: The rapid term.

The model constants used in the present study are given in **Table 2**.

## 3. NUMERICAL APPROACH

The computations for the governing equations can be made using a parabolic marching procedure if the radial pressure gradients are small and the axial diffusion is neglected. Such a situation occurs if velocities in the two streams are comparable. Assuming parabolic conditions, a numerical solver has been developed using finite volume of Patankar [18].

The computations are performed up to an axial distance of approximately 100*D* with an axial forward step size of 0.01 times the local jet half width $\Delta x=0.01L_u$ and 80 grid points in the radial direction are used. The radial expansion $\Delta r=\Delta x\frac{dL_u}{dx}$ is so small that it does not affect the assumption of an orthogonal grid. This means that the grid expands in the radial direction following the jet expansion and this is sufficient to obtain a grid independent numerical solution.

No boundary conditions are prescribed due to the parabolic nature of the flow. The computation progresses from section to section, and its implementation requires only the profiles at the jet nozzle. The boundary conditions at the nozzle exit are those of a fully developed pipe flow [19]. The radial velocity is zero at the nozzle and in the ambient. All variables at the radial jet boundary are equal to those in the ambient. At the axis of symmetry, the radial velocity and the radial gradients of other variables are set to zero. For all calculations, a small co-flow velocity value is used. For the turbulence quantities this implies a value of zero or a negligible small value. The kinetic energy and the Reynolds stress profile are used to derive the energy dissipation through the following relationship:

**Table 2.** Turbulence constants in the second order model, where the value of $C_1$ is adapted for the axisymmetric jet case.

| $C_1$ | $C_2$ | $C_s$ |
|---|---|---|
| 2.3 | 0.6 | 0.22 |

$$\varepsilon=\frac{C_\mu k^2}{\left(-\overline{uv}\right)}\left(\frac{\partial U}{\partial r}\right) \quad (12)$$

## 4. RESULTS AND DISCUSSION

The influence of co-flow is investigated using the experiments data of Djeridane [16]. The jet is ejected from a round nozzle of an internal diameter *D* of 26 mm with various co-flow to jet velocity ratios of $U_{co}/U_j$ = 0.0 and 0.1. The experimental details are given in **Table 3**. The far field behavior of the computed quantities such as the velocity decay constant, the turbulence intensity, the spreading rate and the turbulent flux are presented and discussed.

**Figure 1** shows, the jet centreline axial mean velocity $\left(U_j-U_{co}\right)/\left(\overline{U}_c-U_{co}\right)$ with and without co-flow as a function of the normalized distance *x/D*. It is seen that the velocity is somewhat overpredicted with experiment values. The predicted results obtained, using the two turbulence models, agree quite well with the experimental data of Djeridane [16]. The major visible effect of the co-flow is the jet decay rate reduction, in comparison with the free jet case.

Therefore, the jet without co-flow, tends to mix more rapidly with the ambient air than the co-flowing jets. In the far region, a hyperbolic decrease of the mean velocity is observed. However, for 20 < *x/D* < 50, the velocity decay constant $K_u$, is defined by the following,

$$\frac{\left(U_j-U_{co}\right)}{\left(\overline{U}_c-U_{co}\right)}=K_u\frac{x}{D} \quad (13)$$

**Table 3.** Properties of the investigated turbulent co-flowing jets.

| Jet/co-flow | $U_{jet}$ | $U_{co}$ | $S_\rho$ | $\mathrm{Re}_j$ |
|---|---|---|---|---|
| Air/air | 12 | 0.00 | 1 | 21 000 |
| Air/air | 12 | 1.20 | 1 | 21 000 |

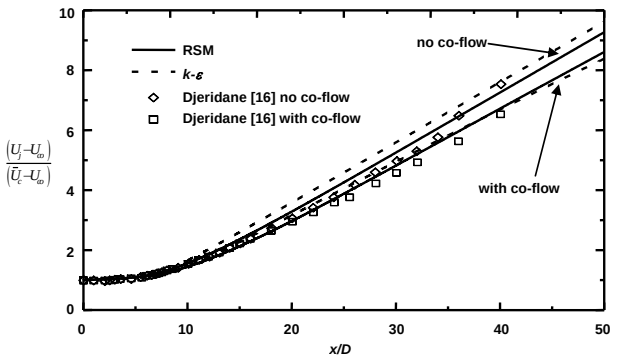



**Figure 1.** Centreline values of the axial mean velocity.

This constant is closely related to the spreading rate, and the asymptotic value of the turbulence intensity.

The comparison of this decay constant with those found in the literature is a good validation tool for the developed computational code. The results of the present investigation give for the case with a co-flow, a value of $K_u$ = 0.188 with the second order model, and a value of $K_u$ = 0.174 with the first order model, while the experimental value is 0.176 Djeridane [16]. However, in the case of the jet without a co-flow, the value for $K_u$ is found to be 0.199 with the second model and 0.201 with the first order model. The experimental value of the slope $K_u$ is 0.218 Djeridane [16]. **Table 4** recapitulates the asymptotic values of the mean velocity decay for various co-flow strengths of the different model predictions and compared to the experimental values of Djeridane [16].

Using the similarity law proposed by Chen and Rodi [20] ($U_j/\overline{U}_c = K_u(x/D_{eq})$), the influence of the velocity ratio on the decay $K_u$ is also described quite well by the most recent experimental study of Wang *et al.* [13] (shows $K_u$ = 0.158, with velocity ratios of $U_{co}/U_j$ = 0.075) and Antoine *et al.* [11] (gives $K_u$ = 0.146, with velocity ratios of $U_{co}/U_j$ = 0.05). The major visible effect of the co-flow, when the velocity ratio increases, is the jet decay rate reduction.

**Figure 2** features a comparison of the computational and experimental jet spreading rates of the velocity field based on the mean velocity half radius $L_u$. It is noticed that the half-width $L_u$ in the jets without a co-flow are much larger than those of the corresponding jets with a co-flow. Furthermore, with the first model, the predicted half-width of the jet obtained by the two jets cases agrees much better with the experimental data of Djeridane [16]. Here, the mean velocity half-width is defined by

**Table 4.** Comparison between the model predictions and measurements of the velocity flow field for asymptotic values of mean velocity decay, the spreading rate and the velocity fluctuation intensities at various co-flow strengths.

| Authors | $U_{co}/U_{jet}$ | $S_\rho$ | $K_u$ | $S_u^{1/2}$ | $\frac{\sqrt{u_c^2}}{(\bar{U}_c - U_{co})}$ | $\frac{\sqrt{u_{\max}^2}}{(\bar{U}_c - U_{co})}$ |
|---|---|---|---|---|---|---|
| RSM | 0.10 | 1.0 | 0.188 | 0.100 | 0.286 | 0.252 |
| RSM | 0.0 | 1.0 | 0.199 | 0.192 | 0.251 | 0.244 |
| *k-ε* | 0.10 | 1.0 | 0.174 | 0.096 | 0.305 | 0.269 |
| *k-ε* | 0.0 | 1.0 | 0.201 | 0.195 | 0.283 | 0.267 |
| Djeridane [16] | 0.10 | 1.0 | 0.176 | 0.132 | 0.261 | 0.219 |
| Djeridane [16] | 0.00 | 1.0 | 0.218 | 0.184 | 0.253 | - |
| Antoine *et al.* [11] | 0.075 | 1. | 0.146 | 0.128 | - | - |
| Wang *et al.* [13] | 0.05 | 1.0 | 0.158 | 0.12 | 0.25-0.30 | - |

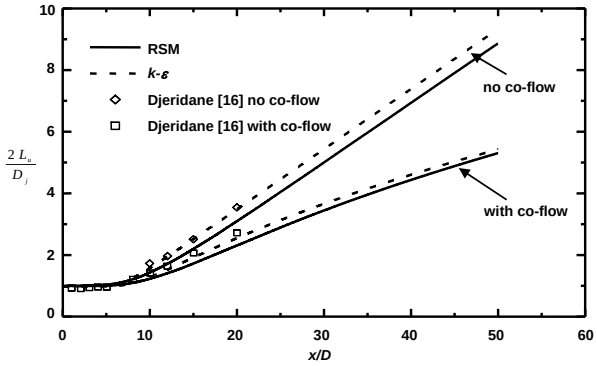



**Figure 2.** Centreline values of the mean velocity halfwidth.

$$\frac{2L_u}{D} = S_u^{1/2}\frac{x}{D} \tag{14}$$

where $S_u^{1/2}$ is the mean velocity spreading rate. In the absence of a co-flow, both models predict a velocity spreading rate $S_u^{1/2}$=0.192 with the second model and $S_u^{1/2}=0.195$ with the first model. These rates are close to the experimental value $S_u^{1/2}=0.184$ of Djeridane [16]. In presence of a co-flow, the half-width is no longer a linear function of *x*, so $S_u^{1/2}$ can not be easily determined. Therefore $S_u^{1/2}$ depends on the axial distance and is not a useful concept in a co-flowing jet. However, the spreading rates $S_u^{1/2}=0.10$ and $S_u^{1/2}=0.096$, obtained by the two models with a co-flow are about 22% smaller than that the average experimental values of Djeridane [16] $S_u^{1/2}=0.132$, the air-air jet of Wang *et al.* [13] $S_u^{1/2}=0.12$, and the water-air jet of Antoine *et al.* [11] $S_u^{1/2}=0.128$. This low value of the velocity spreading rate can be attributed to the presence of the higher co-flowing stream and to the side walls position of the enclosure, compared to those of the latter authors, tending to reduce the jet expansion.

**Figure 3** shows the axial profiles of the velocity fluctuation intensities $\sqrt{u_c^2}/(\overline{U}_c - U_{co})$ on the jet centerline with and without a co-flow. The predicted results obtained by the two turbulence models agree very well with the experimental data of Djeridane [16], and especially with the second order model. The asymptotic value of $\sqrt{u_c^2}/(\overline{U}_c - U_{co})$ is apparently strongly influenced by the used turbulence model. It is seen that the velocity fluctuation intensity with and without a co-flow is slightly overpredicted by the second order model, while for the first order model, the velocity fluctuation intensity is highly overpredicted. This is obvious since the *k-ε* model is an isotropic model which then overestimates the velocity fluctuation intensity. Gharbi *et al.* [21] and Sanders *et al.* [22] have observed this same behaviour and concluded that this deviation is not due to the fact that the *k-ε* model gives unsatisfactory results, rather it is the anisotropy that is badly predicted.

**Figure 3** also features a tendency toward a constant velocity fluctuation intensity value of 0.286 at x/D > 20 with the second order model, and a value of 0.305 with the first order model, which are both close to the value 0.261 obtained by Djeridane [16]. Additionally, it is noticed at x/D < 20 that, without a co-flow, the velocity turbulence intensity increases faster with x/D than for the co-flow case. It is interesting to note that the predicted velocity fluctuation intensity shows an approximate asymptotic behaviour which increases in value with the increasing co-flow velocity.

**Figure 4** shows the radial profiles of the mean velocity for the downstream section *x/D* = 20, situated in the affinity region of the jet. It is noticed that both models agree reasonably well with the experimental data of Djeridane [16] for the co-flow case. **Figure 5** presents the radial velocity fluctuation intensities profiles at *x/D* = 20. Qualitative agreement is obtained in the sense that both models predict a local maximum which is also observed experimentally. It should be mentioned again that the observed difference, between the experimental and the numerical values on the jet axis, is due to the chosen initial conditions. These values are of the order of 15%. The axial mean velocity should decrease faster, and thus

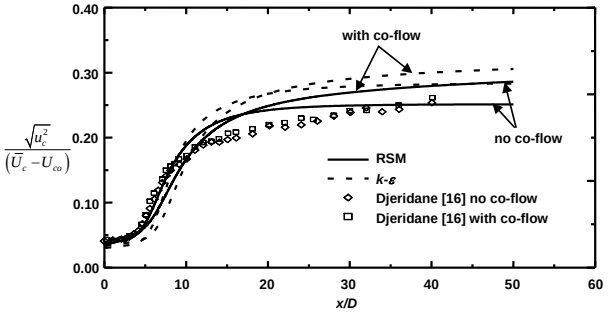



**Figure 3.** Centreline values of the velocity fluctuation intensities.

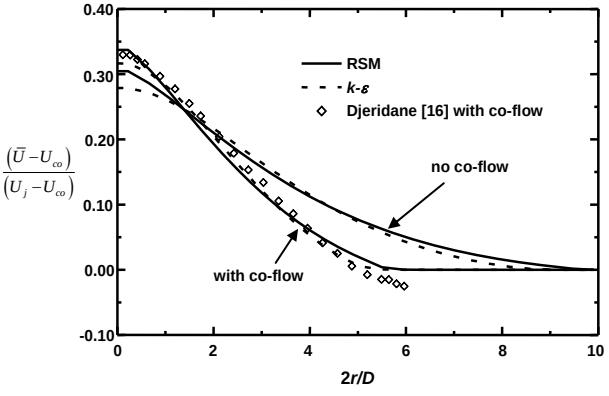



**Figure 4.** Radial profile of the normalized velocity fluctuation intensities at *x/D* = 20.

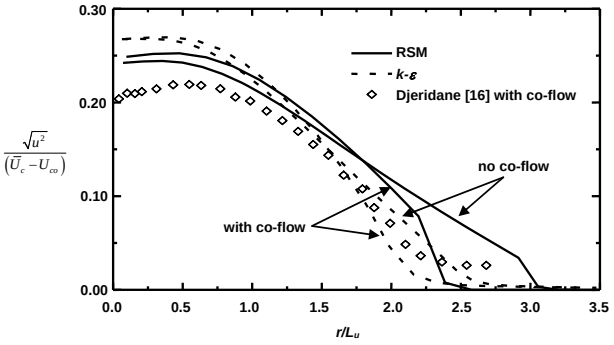



**Figure 5.** Radial profile of the mean velocity at *x/D*=20.

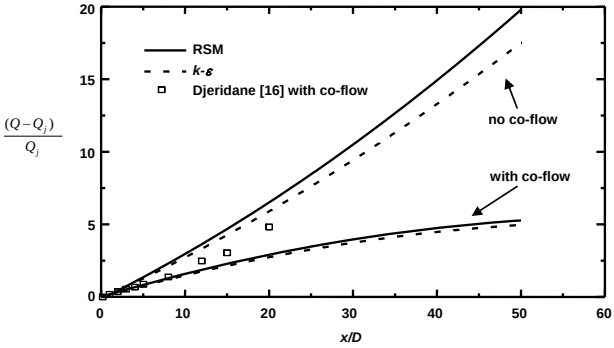



**Figure 6.** Evolution of the axial entrainment of air.

more efficient turbulent mixing is required. Furthermore, it should be noticed that the jet radial expansion is reduced when the co-flow is present.

The amount of air entrainment by the jet is determined by the time-average radial profiles of velocity. It relates the mass flow rate of the surrounding fluid entrained into the jet to the characteristic velocity difference between the jet and the co-flow.

$$Q = 2\pi \int_0^{r(U=Uco)} \rho\left(\overline{U} - U_{co}\right) r\, dr \qquad (15)$$

Based on this last definition, **Figure 6** shows the axial evolution of the air entrainment. It is obvious that the co-flow decreases considerably the air entrainment. A qualitative analysis would suggest that a co-flowing stream would restrict the radial in flow of air into the jet. Moreover, the free jets entrain from 30 to 75% more air than the co-flowing jets at any given axial location.

## 5. CONCLUSIONS

A turbulent jet with and without a co-flowing air has been theoretically and numerically investigated, using the first and the second order turbulence closure models. The calculation results show that both models qualitatively predict the behavior of jets with or without co-flowing air. An investigation of the asymptotic values for

the mean velocity decay constant $K_u$, the spreading rate $S_u^{1/2}$ and the centerline value of the velocity fluctuation intensities has been presented. The predictions agree reasonably well with the very recent experimental study in the literature for axisymmetric jets. The major visible effect of the co-flow is the jet decay rate reduction, in comparison with the free jet case. However, based on entrainment definition, it is mainly shown that the co-flow reduce the air entrainment.

## REFERENCES


[1] Panchapakesan, N.R. and Lumley, J.L. (1993) Turbulence measurements in axisymmetric jets of air and helium, Part 2. Helium jet. *Journal of Fluid Mechanics*, **246**, 225-247.

[2] Ruffin, E., Schiestel, E., Anselmet, F., Amielh, M. and Fulachier, L. (1994) Investigation of characteristic scales in variable density turbulent jets using a second-order model. *Physics of Fluids*, **6(8)**, 2785-2799.

[3] Chassaing, P., Harran, G. and Joly, L. (1994) Density fluctuation correlations in free turbulent binary mixing. *Journal of Fluid Mechanics*, **279**, 239-278.

[4] Lucas, J.F. (1998) Analyse du champ scalaire au sein d'un jet turbulent axisymétrique à densité variable. Ph.D. Thesis, Université d'Aix-Marseille II, Marseille.

[5] Gazzah, M.H., Sassi, M., Sarh, B. and Gökalp, I. (2002) Simulation numérique des jets turbulent subsoniques à masse volumique variable par le modèle *k-ε*. *International Journal of Thermal Sciences*, **41**, 51-62.

[6] Imine, B., Saber-Bendhina, A., Imine, O. and Gazzah, M.H. (2005) Effects of a directed co-flow on a non-reactive turbulent jet with variable density. *International Journal of Heat and Mass Transfer*, **42(1)**, 39-50.

[7] Gazzah, M.H., Belmabrouk, H. and Sassi, M. (2004) A numerical study of the scalar field in turbulent round jet with co-flowing stream. *Computational Mechanics*, **34(5)**, 430-437.

[8] Gazzah, M.H., Belmabrouk, H. and Sassi, M. (2005) Scalar transport modelling in turbulent round jets with co-flowing stream. *International Journal of Thermal Sciences*, **44(8)**, 766-773.

[9] Borean, J.L., Huilier, D. and Burnage, H. (1998) On the effect of a co-flowing stream on the structure of an axisymmetric turbulent jet. *Experimental Thermal and Fluid Science*, **17(1-2)**, 10-17.

[10] Schefer, R.W. and Dibble, R.W. (2001) Mixture fraction field in a turbulent non-reacting propane jet. *American Institute of Aeronautics and Astronautics Journal*, **39(1)**, 64-72.

[11] Antoine, Y., Lemoine, F. and Lebouché, M. (2001) Turbulent transport of a passive scalar in a round jet discharging into a co-flowing stream. *European Journal of Mechanics - B/Fluids*, **20(2)**, 275-301.

[12] Hakem, M., Hazzab, A. and Ghenaim, A. (2007) Experimental investigation of elliptical jet in coflow. *International Journal of Applied Engineering Research*, **2(1)**, 31-43.

[13] Wang, P., Fröhlicha, J., Michelassib, V. and Rodi, W. (2008) Large-eddy simulation of variable-density turbulent axisymmetric jets. *International Journal of Heat and Fluid Flow*, **29(3)**, 654-664.

[14] Curtet, R. (1957) Contribution à l'étude théorique des jets de revolution. *Extrait des Comptes rendus de l'Académie des Sciences*, **244**, 1450-1453.

[15] Steward, F.R. and Gurus, A.G. (1977) Aerodynamic of a confined jet with variable density. *Combustion Science and Technology*, **16(1-2)**, 29-45.

[16] Djeridane, T. (1994) Contribution à l'étude expérimentale de jets turbulents axisymétriques à densité variable. Ph.D. Thesis, Université d'Aix-Marseille II, Marseille.

[17] Schiestel, R. (1993) Modélisation et simulation des écoulements turbulents. Hermès Group, Paris.

[18] Patankar, S.V. (1980) Numerical heat transfer and fluid flow. Hemisphere Publishing, Washington, D.C.

[19] Laufer, J. (1953) The structure of turbulence in fully developed pipe flow. *National Advisory Committee for Aeronautics–Report*–1174, 417-434.

[20] Chen, C.J. and Rodi, W. (1980) Vertical turbulent buoyant jets—a review of experimental data. *The Science and Application of Heat and Mass Transfer*, Pergamon Press, New York.

[21] Gharbi, A., Ruffin, E., Anselmet, F. and Schiestel, R. (1996) Numerical modelling of variable density turbulent jets. *International Journal of Heat and Mass Transfer*, **39(9)**, 1865-1882.

[22] Sanders, J.P.H., Sarh, B. and Gökalp, I. (1997) Variable density effects in axisymmetric isothermal turbulent jets: a comparison between a first-and a second-order turbulence model. *International Journal of Heat and Mass Transfer*, **40(4)**, 823-842.

**Vol.2, No.6, 641-645 (2010)**
doi:10.4236/ns.2010.26080

**Natural Science**

# A simple 2-D interpolation model for analysis of nonlinear data

**Mehdi Zamani**

Department of Civil Engineering, Faculty of Technology and Engineering, Yasouj University, Yasouj, Iran; mahdi@mail.yu.ac.ir

Received 1 March 2010; revised 24 April 2010; accepted 13 May 2010.

## ABSTRACT

**To determination the volume and weight of non-uniform bodies, such as in ore deposits evaluation for mining and rock cutting for construction, the methods of interpolation are usually used. The classic curves, which are frequently used to interpolate one-dimensional data are cubic spline, Bspline and Bezier curves. These methods have good efficiency for determination of geometric characteristics of nonregular masses. They have some limitations and problems with two-dimensional interpolation analysis such as forming large linear systems of equations with a lot of entries and difficulty encounter with their solutions. In this research the two-dimensional splines are used, which have the advantages of simplicity and less computational operations effort. The spline functions that are applied have the continuity of $C^1$ at elements boundaries. The presented model has suitable efficiency for volumes of large extents governing to lots of data.**

**Keywords:** Simulation; Approximation; Least square; Cubic Spline; Optimization; Curve Fitting; Prediction

## 1. INTRODUCTION

The most popular methods for interpolation of data are Lagrange method, Neville iteration approach, Newton divided difference methods, Cubic spline, Hermit spline, Bspline and Bezier methods. For the Neville and Newton divided difference method, the higher degree of polynomial generation is straightforward. In conic spline approach each curve consists of a number of about n segment curves. The general way is to divide the interval to collect subintervals or segments and construct different approximating polynomial on each interval. Approximation by functions of this type is called piecewise polynomial approximation. This method has the advantage of removing the oscillatory nature of high degree polynomials. In this approach there exists the continuity of $C^2$ on the segments boundaries. However; the continuity order is optional to users willing. The formulation for creating cubic spline curve results in the solution of a three-diagonal system of equations. Hermit interpolation are based on two points $p_1$ and $p_2$ and two tangent values $p_1'$ and $p_2'$ at those points. It computes a curve segment that starts at $p_1$, going in direction to $p_1'$ and ends at $p_2$ moving toward $p_2'$. Hermit interpolation has an important advantage. The Hermit curve can be modified by changing the tangent values.

The Bezier curve is a parametric curve $p(t)$ that is a polynomial function of the parameter $t$. The degree of the polynomial depends on the number of points used to define the curve. The method applies control points and presents an approximating curve. The Bezier curve does not pass through internal points but the first and last points. Internal points influence the direction and position of the curve by pulling it toward themselves [1-3]. B-spline methods were first proposed in the 1940 for curves and surfaces. The B-spline curves can be approximating or interpolating curves. The advantage of B-spline curves to Bezier curve is the obtaining the higher continuity for the individual spline segments [2-5]. If there are a set of triple data $c = \{(x_i, y_i, y_i'), i = 0,1,2,\ldots,n\}$ The cubic spline functions $S_i(x)$ can be obtained on each interval $[x_i, x_{i+1}]$. With this model the continuity of $C^1$ exists on the boundary of each segment. The spline function of degree 3 for each interval is:

$$g_i(x) = a_i + b_i x + c_i x^2 + d_i x^3 = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix} \begin{bmatrix} a_i \\ b_i \\ c_i \\ d_i \end{bmatrix} \quad (1)$$

If the **Eq.1** is written based on the first point of each segment $(x_i, x_{i+1})$ then,

$$s_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 =$$
$$\begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix} \begin{bmatrix} 1 & -x_i & x_i^2 & x_i^3 \\ 0 & 1 & -2x_i & 3x_i^2 \\ 0 & 0 & 1 & -3x_i^2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_i \\ b_i \\ c_i \\ d_i \end{bmatrix} \quad (2)$$

The first derivatives of the spline function (2) is as follows,

$$s_i'(x) = b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2 =$$
$$\begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 1 & -2x_i & 3x_i^2 \\ 0 & 2 & -6x_i \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} b_i \\ c_i \\ d_i \end{bmatrix} \quad (3)$$

The coefficients $a_i$, $b_i$, $c_i$ and $d_i$ for each interval can be obtained explicitly from the **Eq.5**.

$$\begin{cases} a_i = s(x_i) = f_i \\ b_i = s'(x_i) = f_i' \\ c_i = \dfrac{3}{h_i^2}(f_{i+1} - f_i) - \dfrac{1}{h_i}(f_{i+1}' + 2f_i') \\ d_i = -\dfrac{2}{h_i^3}(f_{i+1} - f_i) + \dfrac{1}{h_i^2}(f_{i+1}' + f_i') \end{cases}, \quad (4)$$

where $h_i$ is the element size or the element interval ($h_i = x_{i+1} - x_i$). The problem with this approach is that we generally don't have the tangent values at $n + 1$ point. This can be provided by applying the following approximation equations (forward difference, central difference and backward difference approximation, respectively) for the first derivatives at the points and for uniform intervals $h_i$.

$$\begin{cases} f_i' = \dfrac{1}{2h_i}[-3, 4, 1] \\ f_i' = \dfrac{1}{2h_i}[-1, 0, 1], \\ f_i' = \dfrac{1}{2h_i}[1, -4, 3] \end{cases} \quad (5)$$

The following two examples show the interpolation of two sets of data by the above approach.

## 2. EXAMPLES

### 2.1. Example 1

This example is in the field of groundwater engineering. When a production well in unconfined aquifer is pumped, water is continuously withdrawn from storage within the aquifer as the cone of depression progresses radially outward from the well. Because of the absence of a recharge source in the form of vertical leakage, there can be no stabilization of water levels and the head ($h$) in the aquifer will continue to decline supposing the aquifer is infinite in areal extent. However, the rate of decline of head ($h$) continuously decreases as the cone of depression spreads. The partial differential equation governing the unsteady-state radial flow in the nonleaky confined aquifer in polar coordinate system is,

$$\frac{\partial^2 h}{\partial r^2} + \frac{1}{r}\frac{\partial h}{\partial r} = \frac{S}{T}\frac{\partial h}{\partial t}. \quad (6)$$

The Theis solution for the above partial differential equation is,

$$s = h_0 - h = \frac{Q}{4\pi T}\int_u^\infty \frac{e^{-u}}{u} du. \quad (7)$$

Theis approximated the above equation by the following series,

$$h_0 - h = \frac{Q}{4\pi T}\left[-0.5772 - \log_e u + u - \frac{u^2}{2 \cdot 2!} + \frac{u^3}{3 \cdot 3!} - \cdots\right] \quad (8)$$

where,

$$u = \frac{1.87 r^2 S}{Tt}$$

s = drawdown, in feet,
Q = discharge, in gpm,
t = time after pumping started, in days,
T = coefficient of transmisivity, in gallons per day per foot (gpd/ft) and
S = coefficient of storage, dimensionless.

The value of brackets equals $W(u)$. The data for interpolation have logarithm values of $W(u)$ respect to $u$ [6]. For the interpolation of data about ten points are considered. **Figure 1** shows the comparison between the

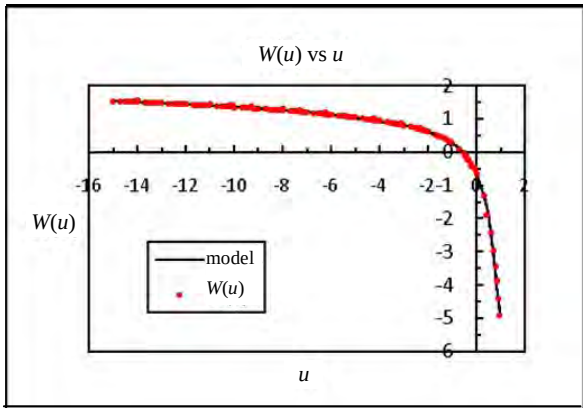



**Figure 1.** The comparison between the data and the model.

data and the model for the nine spline curves. Based on the figure there is a suitable and good relationship between those.

### 2.2. Example 2

In this problem about five pairs of data are obtained from the parametric **Eq.9** with non-uniform segments.

$$\begin{bmatrix} x = t\cos\frac{\pi}{12} - 2\sin\sqrt{2t}\sin\frac{\pi}{12} \\ y = t\sin\frac{\pi}{12} + 2\sin\sqrt{2t}\cos\frac{\pi}{12} \end{bmatrix}, \quad t \in [0,24]. \quad (9)$$

About four cubic spline curves are generated according to the formulations of **Eqs.3**, **4** and **5**. **Figure 2** shows the relationship between the real values and interpolated ones. As it can be seen, there is a relative expectable fitness between them.

## 3. FORMULATION FOR 2-D INTERPOLATION

The bicubic spline formation that applies hear for two dimensional interpolation and four nodes patch is,

$$g_{ij}(x,y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 + a_6 x^3 + a_7 x^2 y + a_8 xy^2 + a_9 y^3 \quad (10)$$

or in matrix notation,

$$g_{ij}(x,y) = [x][A][y] = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix} \begin{bmatrix} a_0 & a_2 & a_5 & a_9 \\ a_1 & a_4 & a_8 & 0 \\ a_3 & a_7 & 0 & 0 \\ a_6 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix}. \quad (11)$$

If the above equation is written based on $(x_i, y_i)$ or the coordinates of the lower left point on each segment, **Eq.12** obtains. Then, the formulation for determination of the coefficient will be simpler.

$$\begin{aligned} s_{ij}(x,y) = {} & a_0 + a_1(x-x_i) + a_2(y-y_j) + a_3(x-x_i)^2 \\ & + a_4(x-x_i)(y-y_j) + a_5(y-y_j)^2 \\ & + a_6(x-x_i)^3 + a_7(x-x_i)^2(y-y_j) \\ & + a_8(x-x_j)(y-y_j)^2 + a_9(y-y_j)^3 \end{aligned} \quad (12)$$

or in matrix notation it is,

$$s_{ij}(x,y) = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix} \begin{bmatrix} 1 & -x_i & x_i^2 & -x_i^3 \\ 0 & 1 & -2x_i & 3x_i^2 \\ 0 & 0 & 1 & -3x_i \\ 0 & 0 & 0 & 1 \end{bmatrix} [A] \begin{bmatrix} 1 & 0 & 0 & 0 \\ -y_i & 1 & 0 & 0 \\ y_i^2 & -2y_i & 1 & 0 \\ -y_i^3 & 3y_i^2 & -3y_i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix}. \quad (13)$$

The above bispline function has ten parameters (unknowns) for each segment or patch. Therefore; ten equations are required for defining them. Four equations satisfy the function values at each corner and six equations satisfying partial derivative in $x$ and $y$ directions for three nodes. Thus there exists the continuity of $C^1$ on each of four sides or boundaries of every element. The first derivatives in $x$ and $y$ direction for the **Eq.12** are:

$$\frac{\partial s_{ij}(x,y)}{\partial x} = \begin{bmatrix} 1 & 2x & 3x^2 \end{bmatrix} \begin{bmatrix} 1 & -2x_i & 3x_i^2 \\ 0 & 1 & -3x_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 & a_4 & a_8 \\ a_3 & a_7 & 0 \\ a_6 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -y_j & 1 & 0 \\ y_j^2 & -2y_j & 1 \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix}. \quad (14)$$

$$\frac{\partial s_{ij}(x,y)}{\partial y} = \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 1 & -x_i & x_i^2 \\ 0 & 1 & -2x_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_2 & a_5 & a_9 \\ a_4 & a_8 & 0 \\ a_7 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2y_j & 1 & 0 \\ 3y_j^2 & -3y_j & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2y \\ 3y^2 \end{bmatrix}. \quad (15)$$

The coefficients $a_0, a_1, \ldots, a_9$ can be derived explicitly from satisfying the following conditions,

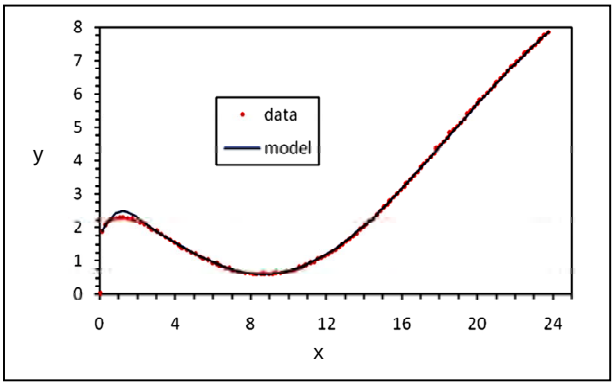



**Figure 2.** The comparison between the data and the model.

$$s_{ij}(i,j)=a_0=f_{ij}\ ,\quad \frac{\partial s_{ij}(i,j)}{\partial x}=a_1=f'_{x,ij}\ ,\quad \frac{\partial s_{ij}(i,j)}{\partial y}=a_2=f'_{y,ij} \tag{16}$$

$$\frac{\partial s_{ij}}{\partial x}(i+1,j)=a_1+2h_ia_3+3h_i^2a_6=f'_{x,i+1,j}\ , \tag{17}$$

$$\frac{\partial s_{ij}}{\partial y}(i,j+1)=a_2+2k_ja_5+3k_j^2a_9=f'_{y,i,j+1}\ , \tag{18}$$

$$\frac{\partial s_{ij}}{\partial y}(i+1,j)=a_2+h_ia_4+h_i^2a_7=f'_{y,i+1,j}\ , \tag{19}$$

$$\frac{\partial s_{ij}}{\partial x}(i,j+1)=a_1+k_ja_4+k_j^2a_8=f'_{x,i,j+1}\ , \tag{20}$$

$$s_{ij}(i+1,j)=s_{i+1,j}(i+1,j)\Rightarrow a_0+h_ia_1+h_i^2a_3+h_i^3a_6+=f_{i+1,j}\ , \tag{21}$$

$$s_{ij}(i,j+1)=s_{i,j+1}(i,j+1)\Rightarrow a_0+k_ja_2+k_j^2a_5+k_j^3a_9+=f_{i,j+1}\ , \tag{22}$$

$$s_{ij}(i+1,j+1)=s_{i+1,j+1}(i+1,j+1)\Rightarrow a_0+h_ia_1+k_ja_2+h_i^3a_3+h_ik_ja_4+k_j^2a_5+h_i^3a_6+h_i^2k_ja_7+h_ik_j^2a_8+k_j^3a_9=f_{i+1,j+1} \tag{23}$$

The above ten equations give a linear system of equations. If it is solved analytically the governing parameters are obtained as follows,

$$a_3=-\frac{3}{h_i^2}a_0-\frac{2}{h_i}a_1+\frac{3}{h_i^2}f_{i+1,j}-\frac{1}{h_i}f'_{x,i+1j}\ , \tag{24}$$

$$a_4=-\frac{1}{h_ik_j}a_0-\frac{1}{k_j}a_1-\frac{1}{h_i}a_2+\frac{1}{h_ik_j}(f_{i+1,j}+f_{i,j+1}-f_{i+1,j+1})+\frac{1}{k_j}f'_{x,ij+1}\ , \tag{25}$$

$$a_5=-\frac{3}{k_j^2}a_0-\frac{2}{k_j}a_2+\frac{3}{k_j^2}f_{i,j+1}-\frac{1}{k_j}f'_{y,ij+1}\ , \tag{26}$$

$$a_6=\frac{2}{h_i^3}a_0+\frac{1}{h_i^2}a_1-\frac{2}{h_i^3}f_{i+1,j}+\frac{1}{h_i^2}f'_{x,i+1j}\ , \tag{27}$$

$$a_7=-\frac{1}{h_i^2k_j}a_0+\frac{1}{h_ik_j}a_1-\frac{1}{h_i^2k_j}(f_{i+1,j}+f_{i,j+1}-f_{i+1,j+1})-\frac{1}{h_ik_j}f'_{x,ij+1}\ , \tag{28}$$

$$a_8=\frac{1}{h_ik_j^2}a_0+\frac{1}{h_ik_j}a_2-\frac{1}{h_ik_j^2}(f_{i+1,j}+f_{i,j+1}-f_{i+1,j+1})-\frac{1}{h_ik_j}f'_{y,i+1j}\ , \tag{29}$$

$$a_9=\frac{2}{k_j^3}a_0+\frac{1}{k_j^2}a_2-\frac{2}{k_j^3}f_{i,j+1}+\frac{1}{k_j^2}f'_{y,ij+1}\ . \tag{30}$$

where $h_i=x_{i+1}-x_i$ , $k_j=y_{j+1}-y_j$,

$$f'_{x,ij+1}=\frac{\partial f}{\partial x}(i,j+1)\ ,\quad f'_{y,i+1j}=\frac{\partial f}{\partial y}(i+1,j)\ . \tag{31}$$

## 4. PROBLEM 3

For testing the above formulation several problems were considered. Here only one of them is presented. The data for this problem are obtained from the **Eq.32**.

$$z(x,y)=4e^{-(0.0867x^2+0.5y^2)}\ ,\quad x,y\in[0,6]\times[0,3] \tag{32}$$

The domain of problem consists of 9 elements and 16 nodes. The elements are uniform where $h_i=2$ and $k_i=1$. The above formulations were applied to to the governing data. **Figure 3** shows the interpolation curves along $x$ = 0, $x$ = 2, $y$ = 0 and $y$ = 2. As it can be seen the 2-D model presents a close relationship to the relating set of data.

## 5. THE RESULTS AND CONCLUSIONS

In this research two models for 1-D and 2-D interpolations were presented. The models are easy and straightforward to handle. Therefore; it can be applied for problems having huge data. It has the advantage of less computational efforts respect to the cubic spline, Bezier and B-spline methods especially for 2-D problems. The continuity for each segment is $C^1$ therefore; this is sufficient for many problems such as calculation of surfaces and volumes of nonuniform bodies. It is recommended to compare the 2-D model with Bezier and B-spine methods respect to the fitness and accuracy.

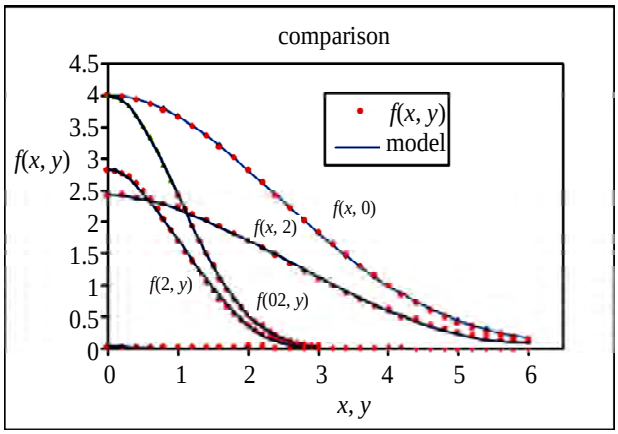



**Figure 3.** The comparison between the data and 2-D model.

## REFERENCES

[1] Salmon, D. (2006) Curves and surfaces for computer graphics. Springer, New York.
[2] Saxena, A. and Sahay, B. (2005) Computer aided engineering design. Springer, Anamaya Publisher, New Delhi.
[3] Conte, S.D. and de Boor, C. (1981) Elementary numerical analysis, an algorithm approach. Kin Keong Printing Co. Pte. Ltd., Singapore.
[4] Burden, R. and Fairs, D. (1989) Numerical analysis. PWS-Kent Publishing Company, Boston.
[5] March, D. (2005) Applied geometry for compute graphics and CAD. 2nd Edition, Springer-Verlag, New York.
[6] Todd, D.K. (1959) Groundwater hydrology. John Wily & Sons, Inc., New York.

**Vol.2, No.6, 646-653 (2010)**
doi:10.4236/ns.2010.26081

**Natural Science**

# Role of the mental foramens in dolphin hearing

**Vyacheslav Ryabov**

Karadag Natural Reserve NAS of Ukraine, Feodosia, Crimea, Ukraine; ryaboff@ukr.net

Received 4 March 2010; revised 16 April 2010; accepted 13 May 2010.

## ABSTRACT

**The role of mental foramens in dolphin hearing was studied in the present work. To this effect the mental foramens' morphology features which are essential from acoustical viewpoint have been studied. The patterns of relationship between the location of mental foramens and their sizes are found. The affinity of the mental foramens' morphology and acoustics that the nature had created testify acoustical expediency of the mental foramens' architecture. This natural inference in the main is confirmed in this work by the experimental data. The mean values of detection thresholds of short broadband stimuli with spectral maximum on frequencies 8, 16, 30 and 100 kHz at acoustical shielding the mental foramens increased on 30, 34, 40 and 50 dB, respectively. Results obtained testify that the mental foramens are the unique sound-conducting pathway into the fat body of the mandibular canal for the sounds of all frequencies used in the experiment, approximately 6−150 kHz (in view of stimuli broadbandness). The left and right row of the mental foramens together with respective mandibular canal plays the role of pinna and external auditory meatus if to use conventional terminology of a land mammal ear. But it is already qualitatively the new external ear implemented by the nature as the receive array and acoustical horn. The new external ear has apparently appeared in result of the dolphins' ancestors' adaptation to new environment conditions, as evolutionary adaptation of the ear to the water and as functional adaptation of the ear in order that to fulfill the new more sophisticated functions in the structure of sonar. The findings give good reason to suppose existence of the same external ear in Odontoceti.**

**Keywords:** Dolphin; Hearing; Architecture of Mental Foramens; Lower Jaw; Unique Pathway; Sound Conduction; The Traveling Wave Antenna; External Ear

## 1. INTRODUCTION

The mechanisms of Odontoceti hearing attract researchers for many years, therefore, a lot of works have been performed in this study area and different hypotheses about sound-conduction mechanisms to a cochlea were suggested in the works. Some researchers assume that the sound passes to the cochlea through the external auditory meatus and the middle ear [1-3]. However, there is an opinion that the auditory meatus cannot participate in sound transmission to the middle ear in general [4-6], or transmits sounds with frequencies under 30 kHz [7,8]. On the grounds of hypothesis related to the sound conduction through the external auditory meatus, authors discuss the possibility of functionally specific subsystems of passive (1-10 kHz) and active (about 100 kHz) hearing [9].

Norris [10] supposed that the sound can be conducted into the fat body of mandibular canal through the mental foramens. Though, little later he suggests different hypothesis [11] which is being generally accepted so far, to the effect that the sound passes into the mandibular fat body directly through the postero-lateral wall of mandible in the area which he called the "acoustic window".

Two "acoustic windows" were determined in electrophysiological experiments: one is for low frequencies (16-22.5 kHz), which is located in the area of external auditory meatus, and the other one - for high frequencies (32-128 kHz) which is located at a distance of 9.3-13.1 cm from melon tip (*i.e.* proximal part of a lower jaw) [12].

It has been also revealed that the fat body conducts the sound to the lateral wall of the tympanic bone, where its thickness is minimum, and the wall play a part of the tympanic membrane, transmitting acoustical vibration to the malleus of the middle ear [6,7,10,11,13-16]. Acoustic stimulation of a lower jaw excites considerable evoked potentials in the dolphin's central auditory system [7]. However, the areas of maximum sensitivity of the mandible surface to sounds of the contact point emitter (size smaller than wavelength) in each work are different, and the test results do not explain the sound-conduction me-

chanisms [6,7,17,18].

Acoustic shielding of the lower jaw significantly hindered the dolphin's ability to discriminate between targets in the echolocation task [19]. Results of these experiments support hypothesis that the lower jaw has a role in the reception of high frequency signals and their transmission to the middle and inner ear. However, the acoustical shield covers the lower jaw from the rostrum tip up to the bases of pectoral fins, *i.e.* it covers both mental foramens, and "acoustic windows", and ventral area of the head, therefore the pathways and mechanisms of sound-conduction so far remain unclear. Along with this, the possibility of simultaneous participation of the auditory meatus and "acoustic windows" for sound conduction to the cochlea at forming spatial auditory image by a dolphin were also discussed [20].

Thus, results presented in the review are inconsistent and the main questions about the pathways and mechanisms of sound-conduction still remain without answers. At the same time the results [21-23] give good reason to consider the mental foramens as the unique sound-conduction pathway into the fat body of dolphin's mandibular canal. The morphology analysis of the dolphin's lower jaw and subsequent modeling confirm this assumption and afford ground to assume that the left and right row of mental foramens from acoustical viewpoint represents the traveling wave antenna which is located in the throat of acoustical catenoidal horn (the left and right mandibular canal plays a role of horn). The model with two traveling wave antennas explains mechanisms of the sound-conduction and directivity of the peripheral part of dolphin's auditory system [21-25]. In agreement with this, anatomical structures of each half of the lower jaw (mental foramens, mandibular canal and fat body) are considered as the components of this part.

We study the role of mental foramens in the dolphin hearing through experiments. The specific tasks of this work are study of the morphology features of mental foramens that are essential from acoustical viewpoint and experimental measurement of the detection thresholds dependence of wide-band acoustical impulses with spectral maximum on different frequencies at mental foramens' acoustical shielding.

## 2. METHODS

### 2.1. Subject and Experimental Conditions

The mandible bones of adult bottlenose dolphin (*Tursiops truncatus p.*) were used for studying the morphology of mental foramens. In order to perform the measurements the mandible was sawn in the area of mental foramens. Further in the text if it was necessary, mental foramens was indicated with the number, as *MFn*, where *n*—is the number of foramen, counting from the rostrum tip; $n = 1, 2, 3, 4$. The cross section dimensions of mental foramens were measured in the plane of maximum width of each foramen, and then in the mutually perpendicular plane.

The experimental studies were carried out with adult Black Sea dolphin (*Tursiops truncatus p.*) which earlier never participated in acoustic experiments, on the base of Karadag Natural Reserve of NAS of Ukraine in the indoor pool of 27.5 × 9.5 × 4.5 m.

### 2.2. Experimental Procedure

The experiments were carried out with using behavioral response techniques (operant conditioning with food reinforcement). A go/no-go response paradigm [26] was used for reporting stimuli condition. The experimental facilities were laid out as follows (**Figure 1**). The dolphin was trained to approach the testing platform (1) upon a trainer's signal, where the trainer was putting on (or was not putting on) the acoustic shield (**Figure 2**) on its rostrum (mental foramens' area). After that the dolphin upon the trainer's signal approached the start ball, and stood touching it by the rostrum tip. Then the researcher switched on the stimulus (or not switched on the stimulus). The stimuli were presented over a period of 4 seconds with repetition frequency 3/second. The dolphin reported that detects the produced stimulus by leaving the start position within 4 s of the trial beginning and touching the signal ball with his rostrum, or if the stimulus was not produced by remaining on the start position within 4 s till a trainer's signal.

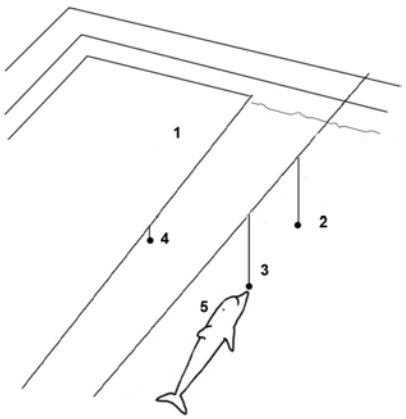



**Figure 1.** The experiment configuration. 1-the testing platform, 2-the stimulus transmitter, 3-the start ball, 4-the signal ball, 5-the dolphin in start position. The dolphin, start ball and stimulus transmitter are located 1 m below the water surface. Signal ball is located near the water surface. The distance between transmitter and pool wall is 3 m, the distance between transmitter and start ball is 2 m.

In these cases the dolphin receives a fish reward for the correct response. The acoustic shield was removed by the trainer every time if it was needed for fish reward. The errors (either dolphin was not approached to the signal ball when signal was produced, or was approached to the signal ball when signal was not produced - false alarm) were not rewarded.

The producing or not producing of stimuli in each trial was determined in random order (but not more than three similar stimuli in sequence). From trial to trial, the stimuli level varied by a one-up-one-down staircase procedure. The session began from a stimuli level well above the anticipated threshold the warming-up part of the session. From trial to trial, the stimuli level varied according to the animal's response in the preceding trial. If the animal detected the stimuli, the level in the next trial decreased one step down. If the animal missed the stimuli, the stimuli level in the next trial increased one step up. The steps were 6 dB in the warming-up part of the session, until the first miss. After that, the steps were 3 dB the measurement part of the session. Responses in stimulus-absent trials did not influence the level in stimulus-present trials. The last one or two trials were again well above the threshold to make sure of ending the session with rewards the cooling-down part of the session. About twenty trials were performed at each step of intensity variation. The threshold was calculated as the mean of all reversal points' maxima and minima of the staircase succession.

## 2.3. Instrumentation

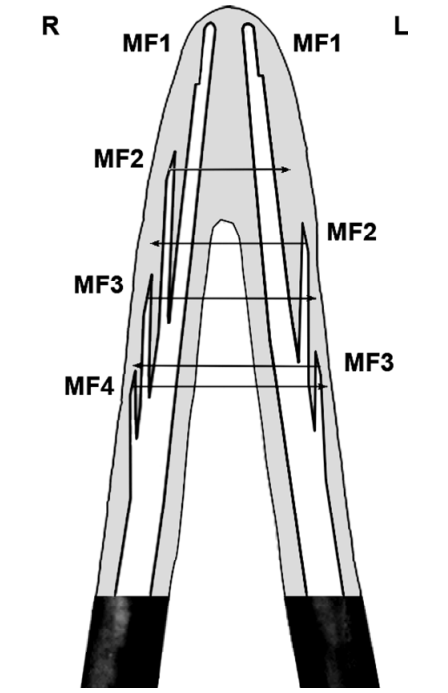
The short broad-band acoustical impulses were used to decrease the effect on experiments results of the direct and reflected signals interference in the pool. Spherical acoustic transducers of 50 or 20 mm diameter were excited by rectangular electrical impulses with the duration of 17 or 5 μs in order to obtain stimuli (the simulated dolphin-like echo-location clicks) with the energy spectrum maximum on frequencies 30 or 100 kHz, respectively. In order to obtain acoustical impulses with the energy maximum on frequencies 8 or 16 kHz, the transducer of 50 mm diameter was excited by rectangular electrical impulse through octave-band filter with center frequency of 8 or 16 kHz, respectively. The duration of each stimulus did not exceed 3 periods of the frequency of its spectrum maximum. In this case the reflections from pool walls and surface of water did not overlap the stimulus, as they were coming with sufficient time delay and were significantly weaker in comparison with the direct signal. Thus, at the stimuli levels near the hearing thresholds, the reflections levels were lower than the detection thresholds of the stimuli. This allows performing the measurements without special sound absorbers.

The acoustically opaque shield of the mental foramens (**Figure 2**) was made as per the shape of dolphin's rostrum so that the shield was put on tightly on the rostrum; the length of the shield was about 15 cm. The shield is produced of the foamed (with closed-cell) neoprene sheet of 5 mm thickness. This waterproof and oil-resistant material has enough durability. Therefore, the efficiency of sound-shielding that gas bubbles of the material provide remains the same for a long time. The acoustical shielding efficiency of this material was tested before the experimentation with using simulated dolphin-like echolocation clicks with spectral maximum on frequencies either 10 or 55 or 170 kHz. Shielding the hydrophone with this material decreased the peak sound pressure level of these dolphin-like clicks up to 28, 32 and 36 dB respectively, for normal incidence of sound.

It is necessary to note, that in frequency range of stimuli used in this experiment the sound wave-length vary from 1.5 to 20 cm, *i.e*. sizes of the acoustic shield on low frequencies are relatively small in comparison with the wave-length. Considering it, the shield covers the upper and lower jaw for providing the best sound shadow in the region of mental foramens (**Figures 2** and **3**). If to make the shield form as external surface of the mandible (*i.e*. shield only lower jaw), then necessary shielding efficiency of the mental foramens, apparently, cannot be received owing to sound diffraction even on the frequency 100 kHz (the sound wave-length 1.5 cm), because distance from edge of the shield up to the mental foramens in this case will be comparable with the wave length (**Figure 3**), let alone low frequencies.

**Figure 2.** The dolphin is near testing platform with the acoustically opaque shield on his rostrum.




(a)


(b)

**Figure 3.** (a) The nasal part (symphysis) of the lower jaw shown in ventral view. In the cross section are presented the layout of MFs of the left (L) and right (R) half of the lower jaw and the region of lower jaw covered by the acoustical shield. The arrows show the projection of external orifices of MFs on the adjacent halves of the lower jaw; (b) The characteristic shape of the oblique endings of mental foramens shown on the left lateral side of the dolphin's rostrum.

## 3. RESULTS

### 3.1. Acoustical Features of the Dolphin's Mental Foramens

The lower jaw's morphology of Odontoceti is similar [11, 27]. The mandible is presented by two rectilinear hollow bones, connected together (in shape of the V or Y letter) in the nasal part along midline of the mandibular symphysis. The cavity of mandible is filled up with the fat body and the corresponding neurovascular bundle. In the area of symphysis the mandibular canal (MC) is pierced by the mental foramens in the places where the branches of mental nerve (n. mentalis) and blood-vessels go out of the mandibular canal on an external surface of mandible.

The cross section size of the neurovascular bundle is much smaller than the sounds wavelengths propagating along the mental foramens and the mandibular canal. The acoustic impedance of tissues filling lower jaw's canals is close to the impedance of sea water [28]. Hence these tissues do not introduce the acoustic heterogeneities and they are transparent for sound. The walls of mental foramens and mandibular canal are acoustically elastic. Using these preconditions, the sound conduction via mental foramens and mandibular canal has been analyzed on the basis of lower jaw's canals geometry [21, 22,23,25].

Let's consider in details the most essential features of structure, shape and sizes of the mental foramens of the dolphin (*Tursiops truncatus p.*) from acoustical viewpoint. The most vivid feature is the amount of the mental foramens. On the left half of the lower jaw there are three of them (**Figure 3**); on the right half - four, what are characteristically not only for the bottlenose dolphins, but for the others species of Odontoceti [11,27].

The mental foramens have frontal direction and oblique ends. The oblique endings of first foramens are somewhat shorter as compared to the others. The distances between the mental foramens are decreasing with the distance from rostrum tip (**Table 1**). It's quite curious

**Table 1.** The mental foramens' basic dimensions of Black Sea dolphin (*Tursiops truncatus p.*). The first and second numbers in column "MF cross section area" are the values from external and internal end of the foramen, respectively. The lengths of MFs are indicated without taking of the oblique ends into account.

| MF number | MF length (mm) | MF cross section area ($mm^2$) | Distance to the next MF (mm) | Length of oblique end (mm) |
|---|---|---|---|---|
| 1 | | | | |
| left | 10 | 8.04; 8.5 | 50.2 | 7 |
| right | 11 | 7.28; 8 | 36.1 | 8 |
| 2 | | | | |
| left | 27.5 | 5; 7.7 | 31.2 | 11 |
| right | 35 | 4.29; 6.04 | 31.5 | 13.5 |
| 3 | | | | |
| left | 12 | 5; 5.7 | | 11 |
| right | 20 | 4.12; 4.94 | 19.15 | 11.5 |
| 4 | | | | |
| right | 10.5 | 1.53; 1.53 | | 10 |

that between the foramens' location and their sizes there are certain patterns of relationship. The mental foramens of the right half of lower jaw is longer than the foramens with the same number of the left half, *i.e*. if the length of mental foramens is denoted as $L_{mn}$, where $m$ is the left ($\ell$) or right ($r$) half of the lower jaw: $m = \ell$ or $r$; $n$ is the index of the mental foramens: $n$ = 1, 2, 3 *or* 4; then, obviously, $L_{r1} > L_{\ell 1}$, $L_{r2} > L_{\ell 2}$ and $L_{r3} > L_{\ell 3}$. Besides, if the foramen is further from the rostrum tip, the length of the foramen (except of MF1) is decreasing, *i.e*. $L_{r2} > L_{\ell 2} > L_{r3} > L_{\ell 3} > L_{r4}$. The length of MF1 in each half of the mandible is essentially shorter than the distance between external orifices of MF1 and MF2. At the same time the length of MF2 and MF3 in the right half of the mandible is few mm longer than the respective distances between external orifices of MF2-MF3 and MF3-MF4 (**Table 1**). Whereas, the length of MF2 in the left half is a few mm shorter than the distance between the orifices of MF2-MF3 (**Figure 3**). In the area of mental foramens the cross-sectional dimensions of the mandibular canal and its cross-sectional area gradually increases (caudal), in respect to each mental foramen. Whereas the cross-sectional dimensions of the mental foramens are decreasing with the distance from rostrum tip, though not as harmoniously as their lengths. Due to this the ratio of cross-sections areas of *MFn/MC* in the area of mental foramens is decreasing (caudal).

But the most obvious and interesting is that the mental foramens of the left and right halves of the lower jaw (except of MF1) are located asymmetrically relatively to longitudinal axis of the animal, though the both halves of the lower jaw are symmetric. Due to the non-equidistant and asymmetrical locations of the mental foramens on both halves of the mandible the projection of each mental foramen (beginning from MF2) on the adjacent half of the lower jaw is located between respective mental foramens (arrows in the **Figure 3**), *i.e*. the location of mental foramens and their lengths are mutually complementally. This vivid result of morphology study obviously shows to us the affinity of acoustics and morphology of mental foramens that we see in their left-right mutually complementary asymmetry.

### 3.2. Influence of Acoustic Shielding of Mental Foramens on Dolphin Hearing

In the present study, we examined the effect of acoustical shielding of the mental foramens on the detection thresholds by the dolphin of the short broad-band sound stimuli with different spectrum maximums of 8, 16, 30 and 100 kHz. The results of measurements are presented in form showing up the relative impairment of the dolphin's hearing at acoustical shielding of the mental foramens in dependence on the frequency of the stimuli spectrum maximum (**Figure 4**). In all tested frequency range (6-150 kHz, taking into account the broadbandness of stimuli), the shielding efficiency is high and increases with the frequency from 30 to 50 dB.

The sound-wavelength in water in dependence on frequency is presented on the same figure for best understanding of the results. The fact that the shielding efficiency dependence on the stimuli frequency is a sort of mirror reflection of the sound-wavelength dependence on frequency (**Figure 4**) proves that the shielding efficiency is inversely proportional to the stimuli wavelength and it means that the shielding efficiency is determined by the stimuli wavelength.

The absolute values of the detection thresholds of stimuli measured in this work agree with the bottlenose dolphin audiogram [29], taking into account phenomenon of the hearing energy summation [30,31]. Therefore, the dolphin, which was used in our experiment, has got normal hearing.

## 4. DISCUSSION

The studying of morphology features of the mental foramens was rather the successful but has created the new question. With what purpose the nature has created such complex harmony of mutually-complementary asymmetric architecture of the mental foramens? For the answer to this question it is necessary to take into account the fact that the left and right row of the mental foramens we considered like the traveling wave antenna and the lower jaw as a peripheral part of a dolphin's hearing system. Therefore we consider the studied morphology features of the mental foramens from viewpoint of the linear arrays' theory and physical acoustics. In this case [21-25], the mental foramens being acoustically narrow waveguides within the range of dolphin's hearing frequencies conduct the sound into the fat body of mandibular canal without distortion and define its intensity. The length of mental foramens and their location defines the delays of sound conduction in the traveling wave antenna. In other words, the mental foramens set the amplitude and phase distribution of particles velocities of the left and right antenna arrays. The oblique ends of the mental foramens, being the waveguides endings, adjust the active acoustical lengths of the mental foramens depending on frequency. The mental foramens are functioning as elementary receivers of arrays and the structure of their location defines the beam pattern of each antenna array. Therefore, mutually-complementary asymmetric architecture of the mental foramens is apparently necessary in order that to form the features of beam patterns of the left and right arrays. The preliminary calculations point out that beam patterns of the left

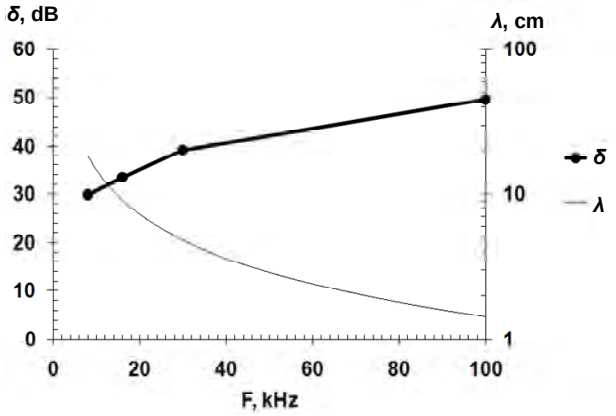



**Figure 4.** ($\delta$) The ratios of mean values of the detection thresholds of stimuli at shielding of the mental foramens to the thresholds without shielding in dependence on the frequency of the stimuli spectrum maximum (F). The sound wavelength in water ($\lambda$) in dependence on frequency (F).

and right traveling wave antennas mutually intercross in ventral direction [21]. Due to this we can assume that in the medial plane of this area the beam patterns shape is mutually-complementary too.

The morphology features of the mental foramens that we considered in this study point out that their architecture is subordinated to the acoustical expediency. This natural inference in the main is confirmed by the experimental data reviewed below.

At shielding the mental foramens the mean values of detection thresholds of stimuli with spectrum maximum on frequencies 8, 16, 30 and 100 kHz increased on 30, 34, 40 and 50 dB, respectively (**Figure 4**). It means that the shielding significantly impairs the dolphin's hearing both on the low frequencies and on the frequencies of dolphin's whistling and in the echolocation frequencies range. This new original results testify that the mental foramens are the unique sound-conducting pathway into the fat body of the mandibular canal in the frequency range at least from 6 kHz and up to 150 kHz (in view of stimuli broadbandness). This fact does not give any chances for the sound conduction along the other hypothetical pathways [1-3,9,11,12,17,18,20].

At the same time the detection thresholds decreases by 20 dB with increase in the wavelength of stimuli. This fact can be explained by increase in penetration of the sound beyond the bounds of the shield as a result of diffraction. Indeed, the sound wavelength significantly increases when the frequency lowers (**Figure 4**), and if on frequencies about 100 kHz the shield length 10 times more of the wavelength, then for the frequencies about 16 kHz the linear dimensions of the shield is becoming comparable to the wavelength and on frequencies of about 8 kHz the dimensions of the shield are even shorter than the wavelength. Therefore, as it follows from the theory and the results (**Figure 4**, 100 kHz), in this waves lengths range, the shielding efficiency is increasing with increase of the ratio of the shield length to the wavelength.

Thus, if to take into account the diffraction effect, so becomes clear, that the shielding efficiency of the mental foramens would be identical if the ratio of the shield dimensions to the wavelength on the measured frequencies was constantly. It is also confirmed by the fact that the detection thresholds is vary in inverse proportion to the wave length, (**Figure 4**). Unfortunately, on the low frequencies it is difficult to perform the same ratio of the shield dimensions to the wavelength, as for frequency of 100 kHz, because the shield dimensions must be too big (for example, 1.875 m, for the frequency of 8 kHz). Therefore, the efficiency of the used shield decreases with lowering frequency.

In view of obtained results the assumption that the morphological structures of lower jaw is the specialized peripheral part of dolphin's auditory system [21-25] that was based on the morphology study and modeling has now the additional experimental confirmation. Indeed, the affinity of the architecture of mental foramens and acoustics which the nature had created (**Figure 3**) apparently points out for us that the left and right row of mental foramens together with respective mandibular canal plays a role of external ear, if to use conventional terminology of a land mammal ear. But it already is qualitatively the new external ear implemented by the nature as the receive array and acoustical horn [21-25]. Moreover, each row of mental foramens being the receive array along with sound-conduction participates in the hearing directivity creation similarly to a land mammal pinna. Whereas, the mandibular canal being the acoustical horn plays a role of both matching device and part of the traveling wave antenna and like an external auditory meatus transmits a sound through a fat body to a tympanic bone's lateral wall [21-25] which plays a role of tympanic membrane [6,7,10,11,13-16].

The new external ear has apparently appeared in result of dolphins' ancestors' adaptation to new environment conditions, as evolutionary adaptation of the ear to the water and as functional adaptation of the ear in order that to fulfill the new more sophisticated functions in the structure of sonar.

Though the shielding efficiency of mental foramens on frequencies lower than 8 kHz was not measured in this work, we can assume that the mental foramens are the unique way for the sound-conduction even on frequencies lower than 8 kHz. The constant inclination steepness of the low-frequency branch of a dolphin's audiogram (9-10 dB/octave) [29] from 50 Hz and up to

20-30 kHz is evidencing in favor of this.

## 5. CONCLUSIONS

It is known that the Odontoceti experienced a number of functional and morphological modifications in the process of secondary adaptation to the aquatic habitat. Apparently, the new functions of mental foramens and mandibular canal in quality of the external ear that were discovered in this work belong to these modifications. The findings of this work give good reason to suppose the existence of the same external ear in Odontoceti.

## 6. ACKNOWLEDGEMENTS

Author would like to thank Sveta Yahno and Nadya Zhukova for the dolphin training. The experimental part of study was supported by the Ukrainian State fund of fundamental researches within the bounds of the joint project “sffr - rfbr - 2009 Ф28.4/024” with Russian Foundation for basic researches.

## REFERENCES

[1] De Haan, F.V.R. (1957) Hearing in whales. *Acta Otolaryngologica Supplementum*, **134**, 1-14.

[2] Fraser, F.C. and Purves, P.E. (1960) Hearing in the cetaceans: Evolution of the accessory air sacs and the structure and function of the outer and middle ear in recent cetaceans. *Bulletin of the British Museum* (*Natural History*), *Zoology*, **7(1)**, 1-140.

[3] Van Heel, W.H.D. (1962) Sound and cetacean. *Netherlands Journal of Sea Research*, **1(4)**, 407-507.

[4] Yamada, M. (1953) Contribution to the anatomy of the organ of hearing of whales. *The Scientific Reports of the Whales Research Institute*, **8(1)**, 1-79.

[5] Fleischer, G. (1973) Structural analysis of the tympanicum complex in the bottlenose dolphin (*Tursiops truncatus*). *Journal of Auditory Research*, **13**, 178-190.

[6] McCormick, J.G., Wever, E.G., Palin, J. and Ridgway, S.H. (1970) Sound conduction in the dolphin ear. *Journal of the Acoustical Society of America*, **48(6)**, 1418-1428.

[7] Bullock, T.H., Grinell, A.D., Ikezono, E., Kameda, K., Katsuki, J., Nomota, M., Sato, O., Suga, N. and Yanagisawa, K. (1968) Electrophysiological studies of central auditory mechanisms in cetaceans. *Zeitschrift für Vergleichende Physiologie*, **59(2)**, 117-156.

[8] Airapetiantz, E.S., Voronov, V.A., Ivanenko, U.V., Ivanov, M.P., Ordovsky, D.L., Popov, V.V., Sergeev, B.F. and Chilingiris, V.I. (1973) The physiology of the sonar system in Black Sea dolphins. *Zhurnal Evoliutsionnoĭ Biokhimii i Fiziologii*, **9(4)**, 416-422.

[9] Dubrovsky, N.A. (1990) On the two auditory subsystems in dolphins: Sensory abilities of cetaceans. Plenum Press, New York, 233-254.

[10] Norris, K.S. (1964) Some problems of echolocation in cetaceans. In: Tavolga, W., Ed., *Marine Bio-Acoustics*, Pergamon Press, New York, 316-336.

[11] Norris, K.S. (1968) The evolution of acoustic mechanisms in odontocete cetaceans. In: Drake, E.T., Ed., *Evolution and Environment*, Yale University Press, New Haven, 297-324.

[12] Popov, V.V., Supin, A.Y., Klishin, V.O., Tarakanov, M.B. and Pletenko, M.G. (2008) Evidence for double acoustic windows in the dolphin, Tursiops truncates. *Journal of the Acoustical Society of America*, **123(1)**, 552-560.

[13] Ketten, D.R. (1997) Structure and function in whale ears. *Bioacoustics*, **8(1-2)**, 103-135.

[14] Nummela, S., Reuter, T., Hemila, S., Holmberg, P. and Paukku, P. (1999) The anatomy of the killer whale middle ear (*Orcinus orca*). *Hearing Research*, **133(1-2)**, 61-70.

[15] Hemilä, S., Nummela, S. and Reuter, T. (1999) A model of the odontocete middle ear. *Hearing Research*, **133(1-2)**, 82-97.

[16] Koopman, H.N., Budge, S.M., Ketten, D.R. and Iverson, S.J. (2006) Topographical distribution of lipids inside the mandibular fat bodies of Odontocetes: Remarkable complexity and consistency. *Journal of Oceanic Engineering*, **31(1)**, 95-106.

[17] Møhl, B., Au, W.W.L., Pawloski, J. and Nachtigall, P.E. (1999) Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region. *Journal of the Acoustical Society of America*, **105(6)**, 3421-3424.

[18] Brill, R.L., Moore, P.W.B., Helweg, D.A. and Dankiewicz, L.A. (2001) Investigating the dolphin’s peripheral hearing system: Acoustic sensitivity about the head and lower jaw. *Technical Report*–1865, 1-14.

[19] Brill, R.L. (1988) The jaw-hearing dolphin: Preliminary behavioral and acoustics evidence. In: Nachtigal, P. and Moore, P., Ed., *Animal Sonar. Processes and Performance*, Plenum Press, New York, 281-287.

[20] Rimskaya-Korsakova, L.K. and Dubrovsky, N.A. (1998) Two sound conduction pathways to the cochlea participate information of spatial auditory image in the dolphin. *Sensory Systems*, **12(4)**, 497-506.

[21] Ryabov, V.A. (2004) Lower jaw–peripheric part of the dolphin echolocation hearing. In: Belkovich, V.M., Ed., *Marine Mammals of the Holarctic*, *Collection of Scientific Papers*, KMK Scientific Press, Moscow, 483-489.

[22] Ryabov, V.A. (2007) Mechanisms of a dolphin’s echolocation hearing. In: Dible, S., Dobbins, P., Flint, J., Harland, E. and Lepper, P., Eds., *Bio-Acoustics* 2007, *Proceedings of the Institute of Acoustics*, **29(3)**, 283-293.

[23] Ryabov, V.A. (2008) Acoustic clutter field and echo reception by the dolphin. *Biophysics*, **53(3)**, 237-242.

[24] Ryabov, V.A. (2003) A dolphin lower jaw is a hydroacoustic antenna of the traveling wave. *Journal of the Acoustical Society of America*, **114(4)**, 2414-2415.

[25] Ryabov, V.A. (2008) Properties of the dolphin’s mandible horn. In: Boltunov, A.N., Ed., *Marine Mammals of the Holarctic*, *Collection of Scientific Papers*, Astroprint Publishing House, Odessa, 468-473.

[26] Schusterman, R.J. (1980) Behavioral methodology in echolocation by marine mammals. In: Busnel, R.-G. and Fish, J.F., Ed., *Animal Sonar Systems*, Plenum Publishing Corporation, New York, 11-41.

[27] Agarkov, G.B., Homenko, B.G. and Hadjinsky, V.G. (1974) Morphology of dolphins. Naukova Dumka, Kiev, 1-167.

[28] Varanasi, U. and Malins, D.C. (1971) Unique lipids of the porpoise (*Tursiops gilli*): Differences in triacyclglyc-

erols and wax esters of acoustic (mandibular and melon) and blubber tissues. *Biochimica et Biophysica Acta*, **231**, 415-418.

[29] Johnson, C.S. (1967) Sound detection thresholds in marine mammals. In: Tavolga, W.N., Ed., *Marine Bio-Acoustics II*, Pergamon Press, Oxford, 247-260.

[30] Johnson, C.S. (1968) Maskes tonal thresholds in the bottlenose porpoise. *Journal of the Acoustical Society of America*, **44(4)**, 965-967.

[31] Babushina, E.S. (1999) Sound reception in marine mammals: Effects of stimulus parameters and transmission pathways. *Biophysics*, **44(6)**, 1064-1071.

**Vol.2, No.6, 654-666 (2010)**
doi:10.4236/ns.2010.26082

**Natural Science**

# Finite element modelling of the pull-apart formation: implication for tectonics of Bengo Co pull-apart basin, southern Tibet

**Ganesh Raj Joshi*, Daigoro Hayashi**

Simulation Tectonics Laboratory, University of the Ryukyus, Okinawa, Japan; *Corresponding Author: ganeshr_joshi@hotmail.com

Received 29 January 2010; revised 12 March 2010; accepted 16 April 2010.

## ABSTRACT

**The tectonic deformation and state of stress are significant parameters to understand the active structure, seismic phenomenon and overall ongoing geodynamic condition of any region. In this paper, we have examined the state of stress and crustal deformation during the formation of the Beng Co pull-apart basins produced by an en-échelon strike-slip fault systems using 2D Finite Element Modelling (FEM) under plane stress condition. The numerical modelling technique used for the experiments is based on FEM which enables us to analyze the static behavior of a real and continues structures. We have used three sets of models to explore how the geometry of model (fault overlap and pre-existing weak shear zone) and applied boundary conditions (pure strike-slip, transpressional and transtensional) influence the development of state of stress and deformation during the formation of pull-apart basins. Modelling results presented here are based on five parameters: 1) distribution, orienttation, and magnitude of maximum ($\sigma_{H\,max}$) and minimum ($\sigma_{H\,max}$) horizontal compressive stress 2) magnitude and orientation of displacement vectors 3) distribution and concentration of strain 4) distribution of fault type and 5) distribution and concentration of maximum shear stress ($\sigma_{H\,max}$) contours. The modelling results demonstrate that the deformation pattern of the en-échelon strike-slip pull-apart formation is mainly dependent on the applied boundary conditions and amount of overlap between two master strike-slip faults. When the amount of overlap of the two master strike-slip faults increases, the surface deformation gets wider and longer but when the overlap between two master strike-slip faults is zero, block rotation observed significantly, and only narrow and small surface deformation obtained. These results imply that overlap between two master strike-slip faults is a significant factor in controlling the shape, size and morphology of the pull-apart basin formation. Results of numerical modelling further show that the pattern of the distribution of maximum shear stress ($\tau_{max}$) contours are prominently depend on the amount of overlap between two master strike-slip faults and applied boundary conditions. In case of more overlap between two masters strike-slip faults, $\tau_{max}$ mainly concentrated at two corners of the master faults and that reduces and finally reaches zero at the centre of the pull-apart basin, whereas in case of no overlap, $\tau_{max}$ largely concentrated at two corners and tips of the master strike-slip faults. These results imply that the distribution and concentration of the maximum shear stress is mainly governed by amount of overlap between the master strike-slip faults in the en-échelon pull-apart formation. Numerical results further highlight that the distribution patterns of the displacement vectors are mostly dependent on the amount of overlap and applied boundary conditions in the en-échelon pull-apart formation.**

**Keywords:** State of Stress; Deformation Regime; Pull-Apart Formation; Numerical Modelling; Southern Tibet

## 1. INTRODUCTION

Pull-apart basins are the prominent feature of topographic depression structures formed as result of crustal extension associated with either right-lateral right-stepping or left-lateral left stepping en-échelon strike-slip fault systems [1,2]. They usually show a rhombic to spindle-shape, and occur at different ranges of size from small sag ponds of few millimeters up to several kilometers

such as the Dead Sea basins [3,4]. The ratio between the length and width of the pull-apart basins mainly varies between 3 and 4 [5], but recorded pull-apart basins from different part of the world show significance differences in their geometry and structural characteristics [5-7]. Several mechanisms have been proposed for the formation of the pull-apart basins (**Figure 1**) but the common types of mechanism are 1) local extension between two en-échelon basement strike-slip fault segments 2) a distributed simple strike-slip shear mechanism and 3) the Riedel shear mechanisms. The relative motion of the crust blocks involve in a pull-apart system can either be parallel or oblique and divided into pure strike-slip, transtensional or transpressional (**Figure 2(1)**). The formation of pull-apart basin geometry exhibits different shapes before, during and after the tectonic deformation as illustrated in **Figure 2(2)**.

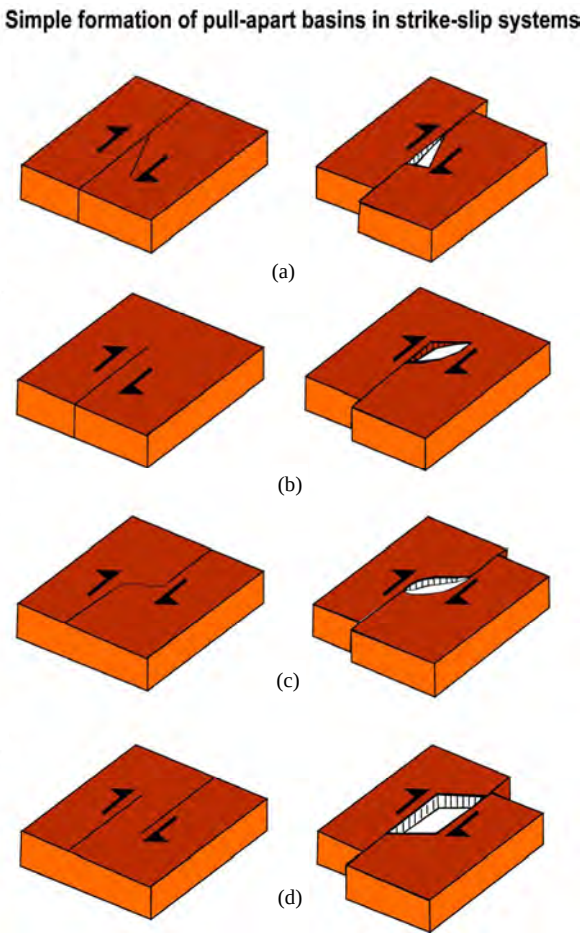



**Figure 1.** Simple formation of pull-apart basins in strike-slip systems. (a) Formation of a pull-apart basin along the releasing band (b) Formation of a pull-apart basin at the termination of a strike-slip fault (c) Formation of a pull-apart basin at the releasing band and (d) Formation of a pull-apart basin at the releasing overstep along a strike-slip fault.

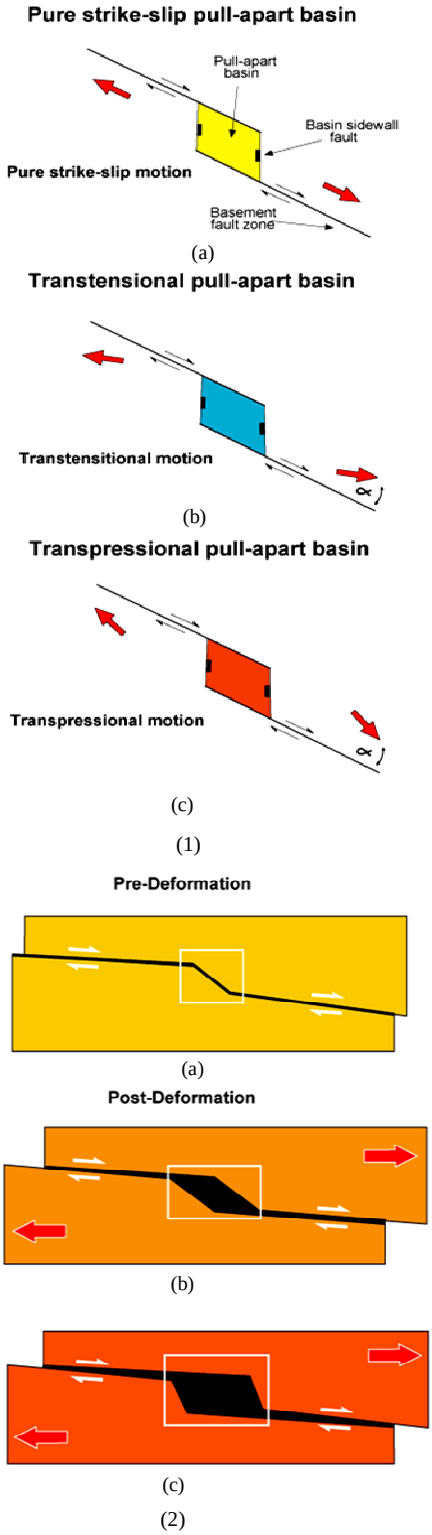



**Figure 2.** (1) General characteristics of strike-slip pull-apart systems developing in (a) pure strike-slip (b) transtensional, and (c) transpressional conditions. (2) Plan view of the formation of pull-apart basin geometry (a) before and (b) during and (c) after tectonic deformation.

Pull-apart basins are the preferred sites of concentrated fracturing [8], elevated heat flow [9] and intense seismicity [10-12]. Moreover, they have significant economical importance and can confine hydrocarbon [13], significant mineralization [14] geothermal fields [15]. Thus, it is important to study the pull-apart basin and their relative role for hydrocarbon aspect. In recent years, many pull-apart basins have been studied extensively in the several parts of the world [11-17]. Several continental pull-apart basins have been also documented in the Tibetan Plateau [11,12] but there is very few studies have been done to focus the pull-apart basin so far. Present study is the first attempt to model numerically Beng Co pull-apart basin in the southern Tibet.

Numerical modelling is a powerful tool, which provides useful insights that are beyond direct observations e.g. stress state, characteristics structures, sequential evolution of the basin, deformation pattern during evolution of the basin, possible temperature regime and rheology during and after the pull-apart formation. Therefore, numerical models have been extensively applied for studying the pull-apart basins [1,2,7,16,18-20]. Segall and Pollard [16] used the analytical models based on the infinitesimal strain theory. They maximized the displaycement near the middle of the faults with the application of remote external stress. These models provide signifycant clues to the orientations of different faults which can develop inside the overstep area. Gölke *et al.* [19] analyzed the vertical displacement and topographic variations in the releasing overstep along the master strike-slip faults by using finite element model. Katzman *et al.* [1] applied the 3D boundary element models of pull-apart basin and compared the modelling results to the Dead Sea Basin. Their results show that the basin deformation mainly depends on the width of the shear zone and on the amount of the overlap between the basin-bounding faults. Petrunin and Sobolev [2,20] presented the 3D thermo-mechanical models of the pull-apart basin developed at an overstepping of an active continental transform faults, and found that the thickness of the brittle layer beneath the basin has significant role in controlling the dimension and deformation pattern of the basin. From their modeling, they further conclude that the deep narrow pull-apart basins are relatively well developed in cold lithosphere, as in the Dead Sea Basin and require very low friction at major faults [2]. Although numerical modelling studies have been applied extensively for simulating deformation in the pull-apart basins, but much less is known overall kinematics or geodynamics within the shallow pull-part structure, as it is filled by unconsolidated sediments, high structurally disrupted or crystallizing materials (veins/plutons).

The purpose of this paper is to understand the relationship between fault geometry; applied boundary conditions (pure strike-slip, transtensional and transpressional), imposed displacements with state of stress and tectonic deformation pattern within a releasing overstep along the two en-échelon strike-slip pull-apart formation applying different sets of models. We have used a series of 2D finite element calculations incorporating elastic rheology under plane stress condition using Mohr-Coulomb failure criterion of faults formation.

## 2. GEOLOGY AND TECTONIC SETTING

The tectonic evolution and uplift of the Tibetan Plateau are a result of tectonic events which occurred due to Indian and Asian plate convergence [21]. The continuing northward movement of the Indian plate for the past 10 Myr has lead to the Tibetan Plateau experiencing widespread extension as indicated by the large scale normal faults and strike-slip zones that made several extensional features such as graben, rift-systems and pull-apart basins in Late Quaternary time [12,21,22]. The tectonic evolution and contemporary states of stress on the Tibetan plateau are mainly governed by E-W extension and N-S compression. The present day average state of stress of the Tibetan Plateau is subject to an extensional ($\sigma_3$) axis trending 112 ± 6° and the minimum horizontal stress ($\sigma_{H\,min}$) trajectory trends WNW-ESE. The compressional ($\sigma_1$) axis trends 022 ± 6° and the maximum horizontal stress ($\sigma_{H\,max}$) trends N-S to NNE-SSW direction, roughly parallel to the Indian-Eurasian convergence in the central part of the India-Asia collision zone [22].

The Beng Co basin is en-échelon strike-slip pull-apart basin named after the 25 km long and 7 km wide Beng Co lake. It developed within the Late Ceneozoic time [12], and is located at 31°10'N and 91°10'E (**Figure 3**). It is about 40 km long with an average strike of north 122°E originating from the long side of Beng Co and extending toward the NW and SE strike-slip fault zone. Geological field observations along the Beng Co can identify, two major fault strands and composed of series of en-échelon pull-apart basins. An en-échelon arrangement of the mole tracks in the field implies possible evidence of the right-lateral strike-slip nature of the Beng Co pull-apart [11]. The Beng Co Fault Zones (BCFZ) cut obliquely across folded Jurassic black shale and calc-schists, whereas the southern branch of the fault zone runs mostly in the granites and the Jurassic shales. The northern exposure of the BCFZ cuts highly folded, early-to-middle Cretaceous red sandstone which lie unconformably upon the Jurassic shales [11]. Further northwest, it passes through the area where ophiolites have been thrust southward on the Jurassic shales and truncates

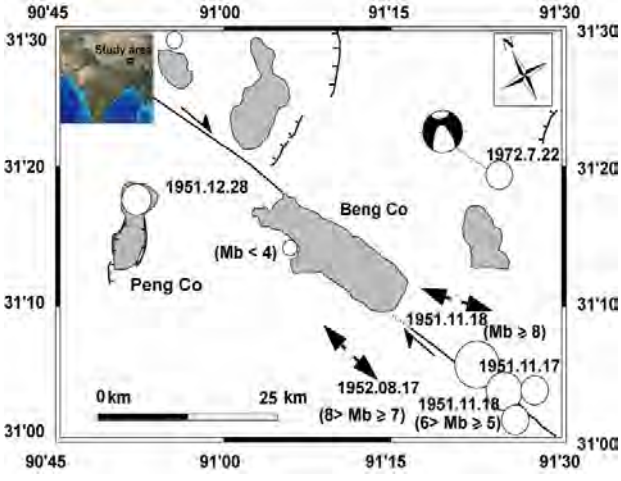



**Figure 3.** Seismotectonic map of the Beng Co region after (Armijo, 1989). Fault plane solution of July 22, 1972, earthquake is from Molnar and Chen (1983) [33]. Black arrows represent tensional directions deduced from analysis of recent minor faults are from Mercier *et al.* (1987) [22].

towards the gently folded conglomerates. The southern branch of the BCFZ lies along the southern edge of a NW-SE granite range.

## 3. SEISMICITY OF THE REGION

The Tibetan Plateau is one of the highest and most active region of the world, which evolved as a consequence of the collision between India and Eurasia landmasses about 50 Ma ago [21]. The continuous northward penetration of Indian crust within Eurasia resulted significant amount of stress accumulation, causing intense seismicity and active tectonic nature of the plateau. In the Tibetan region, seismicity is observed mostly from shallow to intermediate depths. Generally, the seismic pattern shows diffuse in nature and does not follow any known particular tectonic trends. The focal mechanisms solutions here are predominantly of normal and strike-slip type, which further attributed to the large scale E-W extension of the region [23].

The field observations provide several evidences of Quaternary displacements, ruptures and large offsets on either side of the Beng Co pull-apart basin. Several prominent, continuous and fresh surface breaks with large numbers of paleoseismic events along the zone imply that the Beng Co pull-apart region is seismo-tectonically active in contemporary time. Evidence includes several major earthquakes including November 17 and 18, 1951 ($M_W$ = 8); August 17, 1952; December 28, 1951 and July 12, 1972, which show a magnitude ($M_W$) > 6, and are located near the southern extremity of the Beng Co pull-apart (**Figure 3**).

## 4. MODELLING

Numerical simulations are essential for creating an understanding of the physics behind the observations of surface displacement and strain. This is particularly important for understanding data related to active tectonics and earthquake phenomenon because earthquake cycles occur on timescales of thousands of years and our observations sample only a small part of that system. The numerical modelling technique used for the experiments is based on a Finite Element Modelling (FEM) which enables us to analyze the static and behavior of real and continuous structures. FEM has successfully proved to be a powerful method for simulating pull-apart basin geometries and deformation mechanisms, [1,2,7,16,18, 19]. In this study, we applied a 2D-finite element software package developed by Hayashi [24], which has been used widely by Joshi and Hayashi, [25-27]. Similar to most mesh-based numerical methods, bodies of rocks in this program are represented by triangular elements and each element is assigned appropriate material properties, such as density, Young's modulus, cohesion and angle of internal friction. The mesh deforms and moves with respect to material and able to compute appropriate deformation in the program. The details of mathematical formulations about the software package have already described by Hayashi [24].

### 4.1. Model Setup

The dimension of the models are 42 km in length and 7.5 km in width which mimic the natural dimension of the Beng Co pull apart basin adopted after Armijo *et al*. [11] (**Figure 3**). We simplified the model and divide the model area into triangular mesh and several domains. The initial mesh of the model consists of 546 nodal points, 984 triangular elements and two master right-lateral strike-slip faults. In the model, we assumed that the upper crust is a brittle layer and is treated as elastic material. In order to simulate the brittle deformation mechanism of the model, we adopt elastic rheology under plane stress conditions. In our model, the crust up to 20 km is considered to behave as an elastic material because of its brittle nature and presence of earthquake and faults. Rocks forming the brittle crust of the earth contain inhomogeneities which may result in differences compared with our homogeneous and uniform model. In spite of these limitations, our models are still able to yield valuable information related to the pull-apart formation.

### 4.2. Boundary Condition

For the modelling purpose, a two dimension Cartesian rectangular simplified model which shows original ge-

ometry of the Beng Co pull-apart basin has been adopted after Armijo *et al.* [11] (**Figure 4**). Far-field plate velocity boundary conditions are enforced at the either side of the Beng Co Fault Zones (BCFZ). The brittle crust is divided into three simple domains, which may exhibit dissimilar rock layer properties. Domain 1 and 2 represent the southern and northern flank of the pre-existing BCFZ, and domain 3 represents surrounding regions. We consider typical two types of models 1) a model with a pre-existing pull-apart basin and 2) a model without pre-existing pull-apart basin. The model without a pre-existing pull-apart basin is further tested into different overlap/separation ratios (Model B and Model C). We imposed three types of reasonable boundary conditions to mimic the possible natural strike-slip environment of the pull-apart formation. These displacement boundary conditions are 1) pure strike-slip 2) transtensional and 3) transpressional conditions (**Figure 4**). The empirical 100 to 500 m displacements were imposed from northern-left and southern-right corners in different boundary environments, and only 10% of imposed displacement is considered for transtensional and transpressional conditions for modelling (**Figure 4**).

#### 4.2.1. BC1: Pure Strike-Slip Model

The pure en-échelon strike-slip boundary conditions were obtained by moving the upper left-hand and lower-right hand corners using displacement in the left (–X) and right (+X) directions while the lower and upper edges are fixed (**Figure 4(a)**). This boundary condition explores the effect of pure-strike-slip movements on the overall stress field and faulting regime on the pull-apart formation.

#### 4.2.2. BC2: Transtensional Model

The transtensional boundary conditions were simulated by moving the upper left-hand and lower-right hand corners using displacement in the left (–X) and right (+X) directions, and adding an outward displacement in left (–Y) and right (+Y) directions to the lower and upper edges of the model respectively (**Figure 4(b)**). This boundary condition provides the opportunity to understand the distribution and orientation of the stress field and deformation style of the transtensional environment of the pull-apart formation.

#### 4.2.3. BC3: Transpressional Model

In order to investigate the state of stress and overall deformation of the strike-slip pull-apart basin we applied transpressional boundary condition. The transpressional boundary condition were obtained by moving the upper left-hand and lower-right hand corners using displacement in the left (–X) and right (+X) directions with adding an inward displacements in left (–Y) and right (+Y) directions to the lower and upper edges, respectively (**Figure 4(c)**).

### 4.3. Mechanical Parameters and Rock Domain Property

The mechanical properties such as density ($\rho$), Young's modulus ($E$), Poisson's ratio ($\upsilon$), angle of internal friction ($\phi$) and cohesive strength (c) are important rock

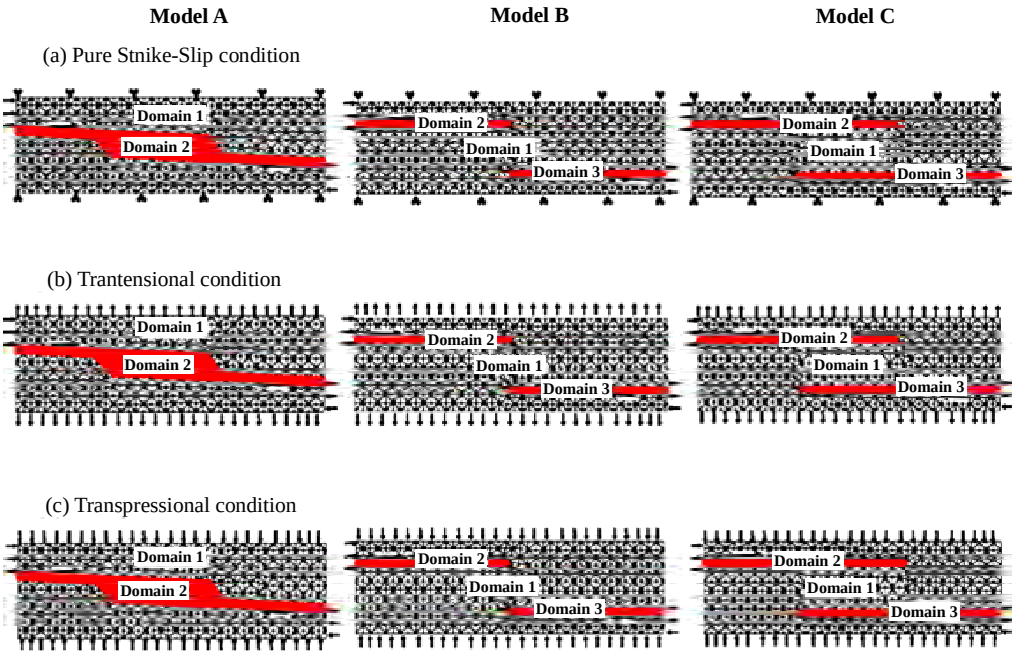



**Figure 4**. Simplified finite element model partition with geometry and boundary conditions for Models A, B and C. The triangular elements show the finite element grid.

parameters in the FEM analysis (**Table 1**). The density ($\rho$) was obtained from the interval velocity of the individual rock domain, using the relation proposed by Barton [28] and compared them with published velocity model [29] from southern Tibet. Seismic *P*-wave ($V_p$) and S-wave ($V_s$) velocities are chosen from the published literature of the study area [30]. We have used two independent elastic constants, Young's modulus and Poisson's ratio to solve the following elastic equations in the brittle part of the lithosphere [24,31].

$$E = \rho V_p^2 \frac{(1+\nu)(1-2\nu)}{(1-\nu)} \quad (1)$$

$$\nu = \frac{1}{2}[1 - \frac{1}{(V_p - V_s)^2 - 1}] \quad (2)$$

where $E$-Young's modulus, $\upsilon$-Poisson's ratio, $\rho$-density of rock, $V_p$- seismic *P*-wave velocity and $V_s$- seismic *S*-wave velocity.

In performing FEM calculation, the whole model is divided into 3 domains and each domain has been allocated distinct rock layer properties on the basis of predominant rock types (**Table 1**). In case of Model A, we assume that BCFZ is pre-existing weak shear zones which allowed us to adopt the value of Young's modulus less compared to other rock domain. ϕ and c were obtained from the Handbook of Physical Constants [32].

## 5. MODELLING RESULTS

To understand the various factors that control the induced state of stress and deformation pattern of the pull-apart basin formation, we have carried out a number of modelling experiments for two characteristic models 1) with a pre-existing pull-apart basin in model and 2) without a pre-existing pull-apart basin in model.

In case of without pre-existing pull-apart basin, we further calculated by two separate models, *i.e.*, Model B and Model C. The Model B which represents no overlap or zero overlap between the two master faults and the Model C corresponds to considerable overlap between two master en-échelon basement strike-slip faults. Each of these models was run for the three most common types of boundary conditions of pull-apart formation: 1) BC1: pure strike-slip condition, 2) BC2: transtensional condition, and 3) BC3: transpressional condition. Here, modelling results are represented based on 1) the maximum ($\sigma_{H\,max}$) and minimum ($\sigma_{H\,min}$) horizontal principle stress 2) magnitude and orientation of the displacement vectors 3) distribution and magnitude of the strain 4) distribution of fault type and 5) concentration and distribution of the maximum shear stress ($\tau_{max}$) contours. The direction and magnitude of the maximum compressive stress axis and minimum compressive stress axis are represented by $\sigma_1$ and $\sigma_3$, respectively. In addition, we have calculated tectonic deformation and faulting regime on the Beng Co pull-apart basin based on the relation and position of the $\sigma_1$, $\sigma_2$ and $\sigma_3$ applying Mohr-Coulomb failure criterion.

### 5.1. Model A: Pre-Existing Pull-Apart Basin

**Figures 5**, **6** and **7** illustrate the orientation of the maximum ($\sigma_{H\,max}$) and minimum ($\sigma_{H\,min}$), horizontal principle stress trajectories, strain distribution, displacement vectors, contour lines of maximum shear stress ($\tau_{max}$) and development of a faulting regime for Model A in the pure strike-slip boundary condition. The calculated $\sigma_{H\,max}$ trajectories show almost N-S directional orientation with uniform distribution in the model with minor variation in the upper left and lower right corners, which corresponds to the direction of maximum shortening of the Tibetan Plateau (**Figures 5(a)**, **6(a)** and **7(a)**). Similarly, $\sigma_{H\,min}$ trajectories show more or less E-W orientation, which is also consistent with the direction of maximum extension in the Tibetan Plateau (**Figures 5(b)**, **6(b)** and **7(b)**). However, some discrepancy was observed in the corners of the models which might be due to boundary effect. The orientations of the displacement vectors show prominent difference among three boundary conditions. The major discrepancy was obtained at the upper-right corner and lower-left corners of the pull-apart basin (**Figures 5(c)**, **6(c)**, **7(c)**). **Figures 5(d)**, **6(d)** and **7(d)** illustrate the predicted strain partitioning for Model A, where high extensional strain is mainly concentrated along pre-existing weak shear zone. This is due to weak rheology, and consistent with the applied model geometry. The predicted faulting pattern shows almost similar predominantly strike-slip type of faults for all boundary conditions (**Figures 5(e)**, **6(e)** and **7(e)**). **Figures 5(f)**, **6(f)** and **7(f)** show concentration and distribution patterns of modeled $\tau_{max}$ contours for all three boundary conditions, where $\tau_{max}$ is largely confined at the central part of the pull-apart basin.

### 5.2. Without Pre-Existing Pull-Apart Basin

In this case, two models (Model B and Model C) were used to calculate state of stress and deformation regime

**Table 1.** Rock mechanical properties used for different domains in the finite element models.

| *Rock Domain* | *ρ (kg/m3)* | *E (GPa)* | *c (MPa)* | *ϕ (deg.)* |
|---|---|---|---|---|
| *Domain* 1 | 2900 | 60.0 | 24.0 | 50.0 |
| *Domain* 2 | 2000 | 01.0 | 10.0 | 31.0 |
| ***Domain* 3** | 2000 | 01.0 | 10.0 | 31.0 |

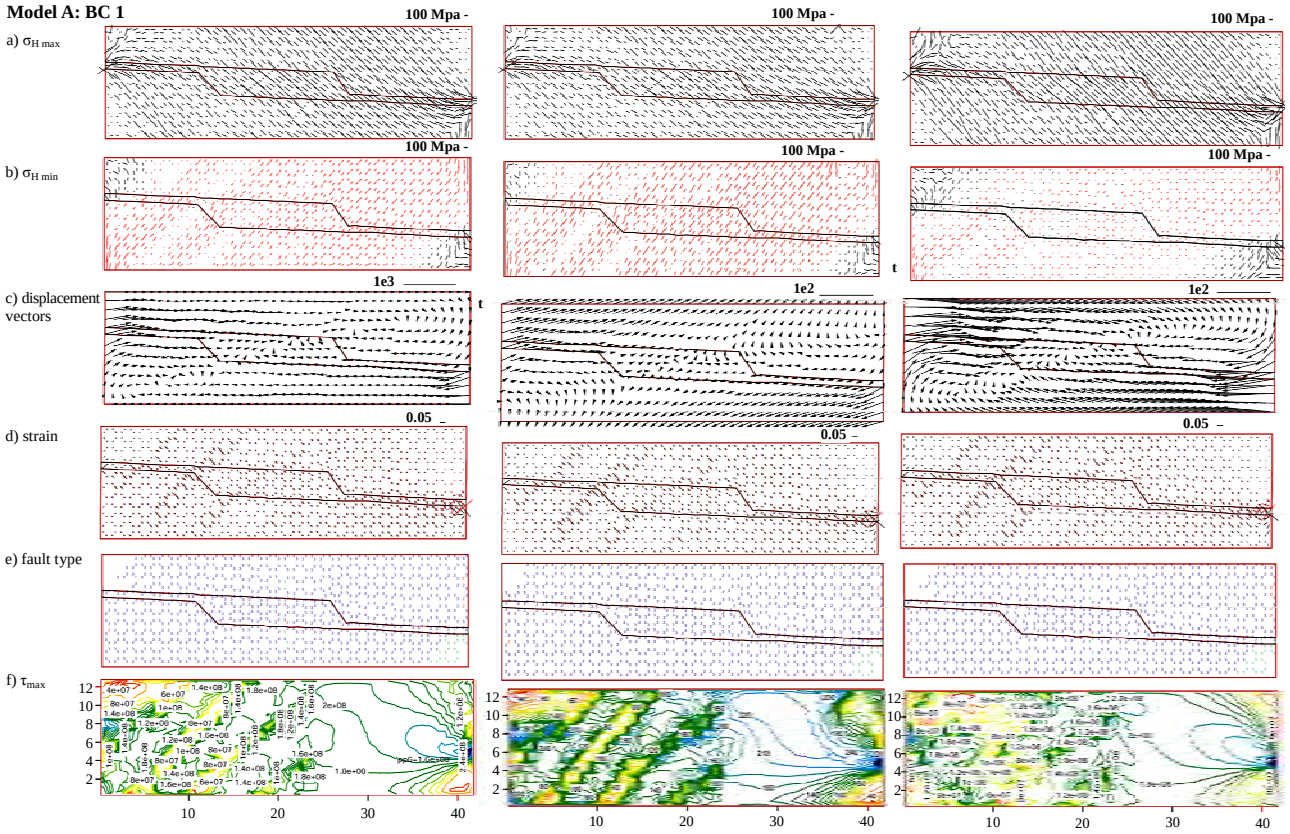



**Figure 5.** Results of Model A for all three pure strike-slip, transtensional and transpressional boundary conditions. (a) Maximum compressional stress ($\sigma_1$) trajectories (b) Maximum extensional stress ($\sigma_3$) trajectories (c) strain distribution (d) displacement vectors (e) faulting regime and (f) distribution of maximum shear stress ($\tau_{max}$) contours under 100 m boundary displacement condition at 10 km depth.

of the Beng Co pull-apart basin for understanding the effect of different overlap on the stress distribution. **Figures 6** and **7** illustrate the calculated maximum ($\sigma_{H\ max}$) and minimum ($\sigma_{H\ min}$) horizontal principle stress trajectories, strain distribution, displacement vectors, contour lines of maximum shear stress ($\tau_{max}$) and development of faulting regime of the Model B and Model C. In both models, orientations of the $\sigma_{H\ min}$ trajectories show more or less E-W directed orientation for all boundary conditions, which is consistent with E-W extension environment of the Tibetan Plateau. A comparison of the Model B and Model C shows that although the general stress ($\sigma_{H\ min}$) patterns remain similar, there are significance differences in the distribution and concentration of $\tau_{max}$ (**Figures 6(f)** and **7(f)**). Similarly, according to applied boundary conditions, the orientation and magnitude of displacement vectors show significant variations between Model B and Model C (**Figures 6(c)** and **7(c)**). There are no considerable differences observed in the predicted strain partitioning among both models, where strain is mainly concentrated along the fault zone which is due to weak rheology. The predicted faulting pattern of the model exhibits almost similar predominantly strike-slip types of faults that have developed for all boundary conditions. If we compare pre-existing pull- apart model (Model A) there is significant difference in distribution and concentration of $\tau_{max}$ contours.

### 5.2.1. Model B: Without Overlap on the Pull-Apart basin

Model B illustrates the results of numerical simulation in the case of no pre-existing pull-apart basin and zero overlap of the two master strike slip faults in the model. **Figure 6** illustrates the orientation of $\sigma_{H\ max}$ and $\sigma_{H\ min}$ trajectories, displacement vectors, strain concentration, distribution of $\tau_{max}$ contours and faulting regimes for Model B. In this model orientation of $\sigma_{H\ max}$ trajectories, strain concentration and faulting regimes which show similar results for all boundary conditions at the same displacement, compared to Models A and C. However, the magnitude of the $\sigma_{H\ min}$ trajectories shows little differences between BC1 and BC3, and the predicted results of displacement vectors and distribution of $\tau_{max}$ show considerable differences between three applied boundary

conditions. **Figure 6(c)** illustrates the principal variations of predicted displacement vectors among three boundary conditions (*i.e.*, BC1, BC2 and BC3) for Model B. Similarly, **Figure 6(f)** shows how differently $\tau_{max}$ is distributed for the different boundary conditions in Model B.

### 5.2.2. Model C: With Fault Overlap on the Pull-Apart Basin

Model C predicted the results of numerical simulation taking into account pre-existing overlap of the two master strike slip faults in the Beng Co pull-apart basin. **Figure 7** illustrates the orientation of $\sigma_{H\,max}$ and $\sigma_{H\,min}$ trajectories, displacement vectors, strain concentration, distribution and accumulation of $\tau_{max}$ and overall faulting regimes for Model C. Results show that there are no considerable variations of the distribution and orientation of the predicted $\sigma_{H\,max}$ and $\sigma_{H\,min}$ trajectories, strain partitioning and faulting regime. Nevertheless, high discrepancies do exist in case of displacement vectors (**Figure 7(c)**) and distribution and concentration of $\tau_{max}$ contours (**Figure 7(f)**). If we compare distribution and concentration of $\tau_{max}$ to other models the Model C does not predict $\tau_{max}$ in the centre of the pull-apart basin which is possibly due to the fault overlap geometry. Moreover, major difference appear in predicted the maximum extensional stress ($\sigma_3$) trajectories within the Model C (**Figure 7(b)**), which might be the cause of the applied boundary condition.

## 6. DISCUSSIONS

### 6.1. Effect of Pre-Existing Weak Shear Zone of Pull-Apart Basin

We first explore the effect of a pre-existing weak shear zone of pull-apart basin on the stress field and deformation pattern during formation of the pull-apart basin. **Figure 5** illustrates the modelling results of a pre-existing weak shear zone of Beng Co strike-slip pull-apart basin. In order to quantify the relative importance of a pre-existing strike-slip weak shear zone on the pull-apart basin the modelling results are compared between Model A and Model B. A close examination of results of

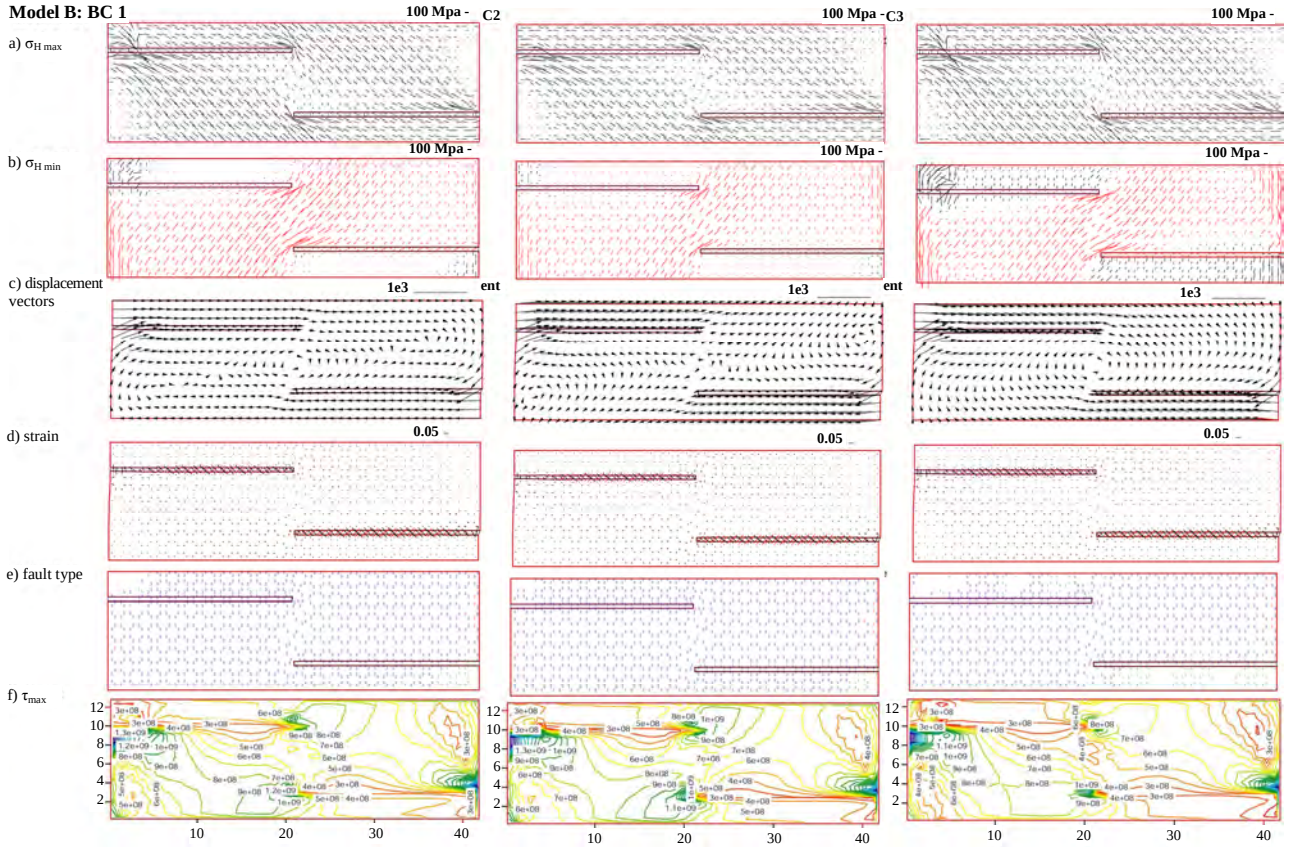



**Figure 6**. Results of Model B for all three pure strike-slip, transtensional and transpressional boundary conditions.(a) Maximum compressional stress ($\sigma_1$) trajectories (b) Maximum extensional stress ($\sigma_3$) trajectories (c) strain distribution (d) displacement vectors (e) faulting regime and (f) distribution of maximum shear stress ($\tau_{max}$) contours under 100 m boundary displacement condition at 10 km depth.

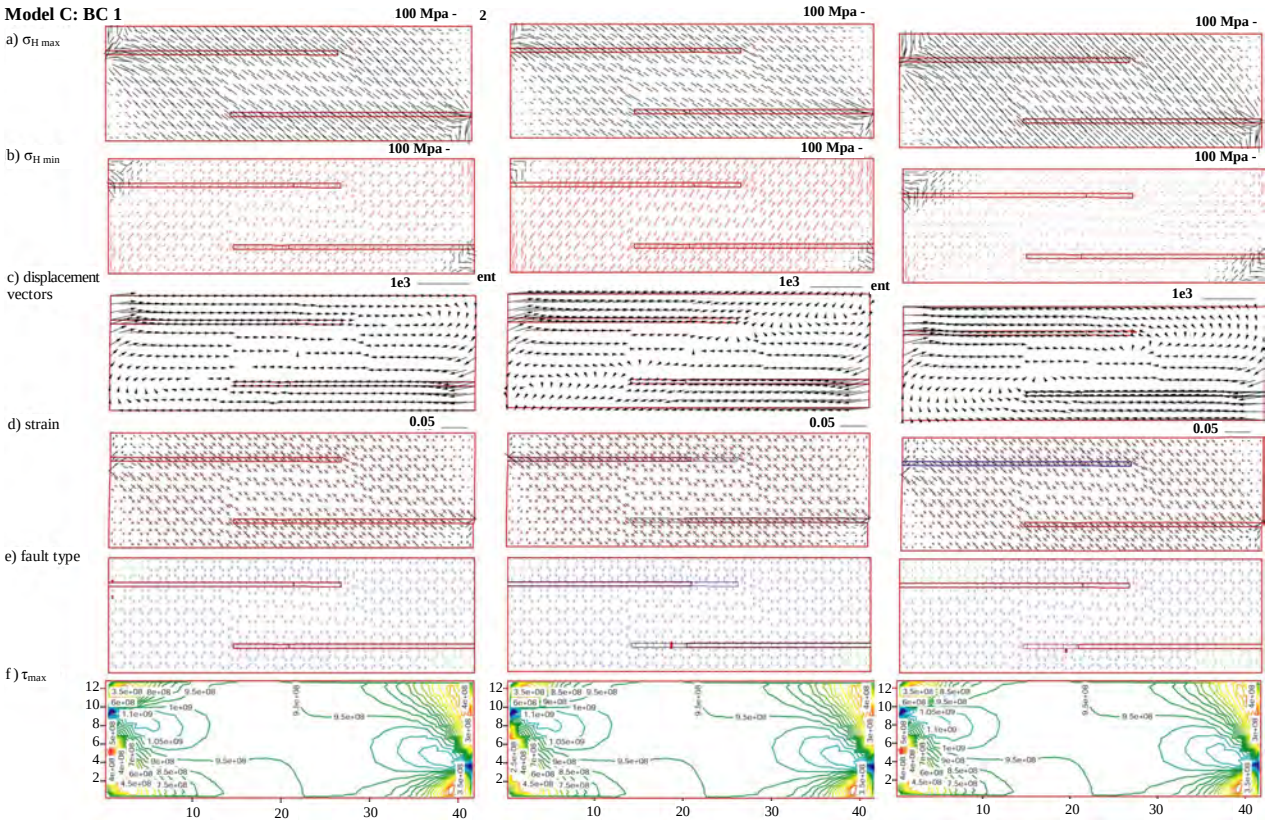



**Figure 7**. Results of Model C for all three pure strike-slip, transtensional and transpressional boundary conditions. (a) Maximum compressional stress ($\sigma_1$) trajectories (b) Maximum extensional stress ($\sigma_3$) trajectories (c) strain distribution (d) displacement vectors (e) faulting regime and (f) distribution of maximum shear stress ($\tau_{max}$) contours under 100 m boundary displacement condition at 10 km depth.

these two models demonstrate that major disparities exist in the horizontal displacement vectors and distribution and concentration of $\tau_{max}$ contour lines, whereas minor differences also exist with regards to the orientation and magnitude of the horizontal principal stresses and deformation pattern, which indicate that the effect of the pre-existing weak shear zone of the pull-apart basin are important control on distribution and concentration of $\tau_{max}$, and principal stresses direction.

## 6.2. Effect of Change in Boundary Conditions in Pull-Apart Formation

Boundary conditions are important factors for controlling the stress state and deformation patterns of the model. Therefore, we explore the effect of a change in boundary conditions on the stress field and deformation style in the formation of the pull-apart basin. In order to investigate the effect of boundary conditions in stress field and deformation patterns, we have tested three types 1) pure strike-slip 2) transtensional and 3) transpressional of boundary conditions. **Figures 5**, **6** and **7** show predicted modelling results of the $\sigma_{H\ max}$, $\sigma_{H\ min}$, $\tau_{max}$, displacement vector, strain partitioned and faulting regime for all three models. Modelling results clearly demonstrated that the distribution and concentration of $\tau_{max}$, displacement vectors and dimension of the pull-apart basin in each boundary condition varies significantly, while orientations of the $\sigma_{H\ max}$ and $\sigma_{H\ min}$ are moderately influenced and faulting regime is not effected by changing applied boundary conditions.

## 6.3. Effect of Change in Fault Overlap in Pull-Apart Development

To investigate the effect of change in the two en-échelon faults overlap geometry we have considered two separate models having 1) zero fault overlap (Model B), and 2) with fault overlap (Model C). **Figure 7** shows the predicted result of fault overlap Model C. If we compare the predicted results of this model with other two models (Model A and Model C) (**Figures 6** and **7**) we observed that the major differences among models are in the orientation of displacement vectors, and distribution and concentration of the $\tau_{max}$ contours. The large rotation of

the horizontal displacement vector appears in the central part of the pull-apart basin with zero overlap model (Model B), while no significant rotation of displacement vector observed in the overlap model (Model C). The Model C produced a tentative rectangular and wide pull-apart basin, while Model B produced a narrow and small pull-apart basin (**Figures 6** and **7**). Moreover, simulated results from our models clearly show that if the faults overlap increases, the size of the pull-apart basin also increases and if the fault overlap decreases the size of the pull apart basin decreases, considerably. These results of numerical modelling imply that fault overlap geometry has an extensive control on the change in shape, size and morphology of the pull-apart formation, which is consistent with previous studies such as Gölke, et al., [19]. Moreover, fault overlap geometry has significant effect on distribution and orientation of $\sigma_{H\,min}$ and concentration of the $\tau_{max}$ contours but there is no effect on the development of fault type (**Figures 6** and **7**).

### 6.4. Effect of Change in Displacement in Pull-Apart Formation

The applied displacement is another significant factor that strongly influences on the magnitude and orientation of the stress field and deformation pattern. We have investigated the effect of applied displacement on the deformation and stress regime during the pull-apart development. We have used 100 to 500 m displacement conditions from the either sides of the model. Our modelling results clearly show that displacement has a major effect on the magnitude and orientation of the maximum ($\sigma_{H\,max}$) and minimum ($\sigma_{H\,min}$) horizontal stresses and displacement vectors, but minor effect on the style of faulting. This result indicates that the change in displacement significantly influences the magnitude of the stress trajectory but only has a limited effect on the orientation of the pull-apart formation. We have further explored the influence of change in displacement on maximum shear stress ($\tau_{max}$) concentration. The model results demonstrate that if we increase the applied displacement the magnitude and concentration of the $\tau_{max}$ contour increases considerably and shear strain will become concentrated in the two ends of the master fault zones.

## 7. CONCLUSIONS

A two-dimension finite element numerical model was used to simulate the strike-slip pull-apart basin formation. We examine the state of stress and deformation associated with the right-lateral, en-échelon Beng Co pull-apart basin in the southern part of Tibetan Plateau. In this paper, we have considered three models each incorporating three different boundary conditions (pure strike-slip, transtensional and transpressional) with different amount of fault overlap of the master strike-slip fault systems. Our modelling results demonstrate that the deformation pattern of the en-échelon strike-slip pull-apart formation is mainly dependent on the geometry of the pull-apart basin, applied boundary conditions and the amount of overlap between two master strike-slip fault systems. When the amount of overlap of the shear zone increases, the surface deformation gets wider and longer between two master faults, but if zero overlap exists between the two strike-slip fault systems, the narrow pull apart formed and block rotation is observed within the pull-apart basin. Based on present modelling we conclude that overlap between two en-échelon strike-slip faults is a significant factor in controlling the shape, size and morphology of the pull-apart formation.

The pattern of the rotation of displacement vectors and maximum shear stress ($\tau_{max}$) distribution contours are also highly dependent on the applied boundary conditions and amount of overlap. In the case of a larger overlap, $\tau_{max}$ is mainly concentrated at two corners of the master strike-slip faults and reduces toward the centre of the pull-apart basin, whereas for zero overlap conditions, $\tau_{max}$ is largely concentrated at the two corners and tips of the master strike-slip faults. These results imply that the concentration and distribution of the maximum shear stress ($\tau_{max}$) is principally governed by amount of overlap between the master strike-slip faults in the en-échelon pull-apart formation.

Finally, on the basis of our modelling results we can conclude that the adopted geometry, applied boundary conditions and amount of overlap of the shear zone have a remarkable role in controlling the overall dimension, stress distribution and deformation pattern during the pull-apart formation.

## 8. ACKNOWLEDGEMENTS

G. R. Joshi gratefully acknowledges the Ministry of Education, Sports and Culture (Monbukagakusho) Japan for the financial support to accomplish this research. The authors wish to thank simulation tectonics laboratory members for their help and support during the research.

## REFERENCES

[1] Katzman, R., Brink, U.S. and Lin, J. (1995) Three dimension of modelling of pull-apart basins: Implications for the tectonics of the Death Sea Basins. *Journal Geophysical Research*, **100(B4)**, 6295-6312.

[2] Petrunin, A. and Sobolev, S.V. (2006) What controls the thickness and lithospheric deformation at a pull-apart basin? *Geology*, **34(5)**, 389-392.

[3] Burchfiel, B.C. and Stewary, J.H. (1966) Pull-apart origin of the central segment of the Death valley. *Geological Society of America*, **77(4)**, 439-442.

[4] Sylvester, A.G. (1988) Strike-slip faults. *Bulletin of Geological Society of America*, **100(11)**, 1666-1703.
[5] Ayden, A.A. and Nur, A. (1982) Evolation of pull-apart basins and their scale independence. *Tectonics*, **1(1)**, 91-105.
[6] Gamond, J.F. (1983) Displacement feature associated with fault zone: A comparison between observed examples and experimental models. *Journal of Structural Geology*, **5(1)**, 33-45.
[7] Bahat, D. (1983) New aspects of rhomb structures. *Journal of Structural Geology*, **5(6)**, 591-601.
[8] Connolly, P. and Cosgrove, J. (1999) Prediction of fracture-induced permeability and fluid flow in the crust using experimental stress data. *AAPG Bulletin*, **83(5)**, 757-777.
[9] Hu, S., O'Sullivian, B.P., Raza, A. and Cona, B.P. (2001) Thermal History and tectonic subsidence of the Bohai Basin, northern China: A Cenozoic rifting and pull-apart basins. *Physics of the Earth and Planetary Interiors*, **126(3-4)**, 121-135.
[10] Armijo, R., Meyer, B., Navarro, A., King, G. and Barkar, A. (2002) Asymmetric slip partitioning in the Sea of Marmara pull-apart: A clue to propagation processes of the North Anatolian Fault. *Terra Nova*, **14(2)**, 80-86.
[11] Armijo, R., Tapponnier, P. and Tonglin, H. (1989) Late Cenozoic right-lateral strike-slip faulting in southern Tibet. *Journal of Geophysical Research*, **94(B3)**, 2787-2938.
[12] Armijo, R., Tapponnier, P., Mercier, L. and Tonglin, H. (1986) Quaternary extension in southern Tibet: Field observations and tectonic implications. *Journal of Geophysical Research*, **91(B14)**, 13803-13872.
[13] Harding, T.P. (1990) Identification of wrench faults using sub-surface structural data: Criteria and pitfalls. *AAPG Bulletin*, **74(10)**, 1590-1609.
[14] Recherds, P.D., Boyce, A.J. and Pringle M.S. (2001) Geological evolution of the Escondida area, northern Chile: a model for spatial and temporal localization of porphyry Cu mineralization. *Economic Geology*, **96(2)**, 271-305.
[15] Monastero, F.C., Katzenstein, A.M., Miller, J.S., Unruh, J. R., Adams, M.C. and Richerds-Dinger, K. (2005) The Coso geothermal field: A nascent metamorphic core complex. *Bulletin of Geological Society of America*, **117(11-12)**, 1534- 1553.
[16] Segall, P. and Pollard, D.O. (1980) Mechanics of discontinuous faults. *Journal of Geophysical Research*, **85(B8)**, 4337-4350.
[17] Basile, C. and Brun, J.P. (1999) Transtensional faulting pattern from pull-apart basin to continental margins: An experimental investigation. *Journal of Structural Geology*, **21(1)**, 23-37.
[18] Du, Y. and Aydin, A. (1993) The maximum distortion energy density criterion for shear fracture propagation with applications to the growth paths of en-échelon faults. *Geophysical research Letters*, **20(11)**, 1091-1094.
[19] Gölke, M., Cloetingh, S. and Fuch, K. (1994) Finite element modelling of pull-apart formation. *Tectanophysics*, **240(1-4)**, 45-57.
[20] Petrunin, A. and Sobolev, S.V. (2008) Three-dimensional numerical models of the evolution of pull-apart basins. *Physics of Earth and Planatery Interiors*. **171(1-4)**, 387-399.
[21] Molnar, P. and Tapponier, P. (1975) Cenozoic tectonics of Asia: Effects of a continental collision. *Science*, **189(4201)**, 419-426.
[22] Mercier, J.L., Armijo, R., Tapponinier, P., Carey-Gailhardis, E. and Han, T.L. (1987) Change from late tertiary compression to late quaternary extension, in southern Tibet, during the India-Asia collision. *Tectonics*, **6(3)**, 275-304.
[23] Torre, T.L. de la, Monsalve, G., Sheehan, A.F., Sapkota, S. and Wu, F. (2007) Earthquake processes of the Himalayan collision zone in eastern Nepal and the southern Tibetan Plateau. *Geophysical Journal International*, **171(2)**, 718-738.
[24] Hayashi, D. (2008) Theoretical basis of FE simulation software package. *Bulletin of the Faculty of Science*, University of the Ryukyus, **85**, 81-95.
[25] Joshi, G.R. and Hayashi, D. (2008a) Neotectonic deformation and shortening along the Himalayan front in the Garhwal region by finite element modelling. *Bullettino di Geofisica Teorica ed Applicacate*, **49**, 228-233.
[26] Joshi, G.R. and Hayashi, D. (2008b) Numerical modelling of neotectonic movements and state of stresses in the central seismic gap region, Garhwal Himalaya. *Journal of Mountain Science*, **5(4)**, 279-298.
[27] Joshi, G.R. and Hayashi, D. (2010) Development extensional stresses in the compressional setting of the Himalayan thrust wedge: Inference from numerical modelling. Natural Science (in press).
[28] Barton, P.J. (1986) The relationship between the seismic velocity and density in the continental crust—a useful constraint? *Geophysics Journal of the Royal Astronomical Society*, **87(1)**, 195-208.
[29] Zhao, W., Nelson, K.D. and Project INDEPTH Team (1993) Deep seismic reflections evidence for continental underthrusting beneath south Tibet. *Nature*, **366(6455)**, 557-559.
[30] Cogan, M.J., Nelson, K.D., Kidd, W.S.F., Wu, C. and Project INDEPTH Team (1998). Shallow structure of the Yadong-Gulu rift, southern Tibet, from refraction analysis of Project INDEPTH common midpoint data. Tectonics, **17(1)**, 46-61.
[31] Timosenko, S.P. and Goodier, J.N. (1970) Theory of elasticity. 3rd Edition, McGraw-Hill Book Company, London.
[32] Clark, Jr., S.P. (Ed.) (1966) Handbook of Physical Constants. New York, Geological Society America, Memoir.
[33] Molnar, P. and Chen, W.P. (1983) Focal depths and fault plane solutions of earthquakes under the Tibetan Plateau. *Journal of Geophysical Research*, **88(B2)**, 1180-1196.

## Appendix

Appendix A is quoted from "Theoretical basis of FE simulation software package" page 84 to 89 written by Hayashi (2008).

### 1. 2D Elastic Problem

The principle of virtual work is described that the external works done by virtual displacement equals the internal work done by virtual strain. Let us consider a certain element within a domain concerned as shown in **Figure A1**. When small displacement $\bar{\mathbf{u}}_i$, which is called virtual displacement, is applied to deform the element without disturb the balance of system, the external work is written as

$$W = \left(\bar{\mathbf{u}}^e\right)^T \mathbf{f}^e \quad \text{where} \quad \bar{\mathbf{u}}^e = \begin{pmatrix} \bar{\mathbf{u}}_1 \\ \bar{\mathbf{u}}_2 \\ \bar{\mathbf{u}}_3 \end{pmatrix} \quad \text{and} \quad \mathbf{f}_e = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix}$$

While taking $\bar{\mathbf{e}}$ as virtual strain derived from virtual displacement and **s** as stress, the strain energy of the element are shown as

$$U = \int_S \left(\bar{\mathbf{e}}\right)^T \mathbf{s} dS$$

According to the principle of virtual work, both must be equated. $\boldsymbol{W} = \boldsymbol{U}$

$$\left(\bar{\mathbf{e}}^e\right)^T \mathbf{f}^e = \int_S \left(\bar{\mathbf{e}}\right)^T \mathbf{s} dS \tag{1}$$

$$\begin{aligned} u &= a_0 + a_i x_i \\ &= a_0 + a_1 x_1 + a_2 x_2 \\ &= (1 \;\; x_1 \;\; x_2)\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \\ &= \mathbf{ca} \end{aligned}$$

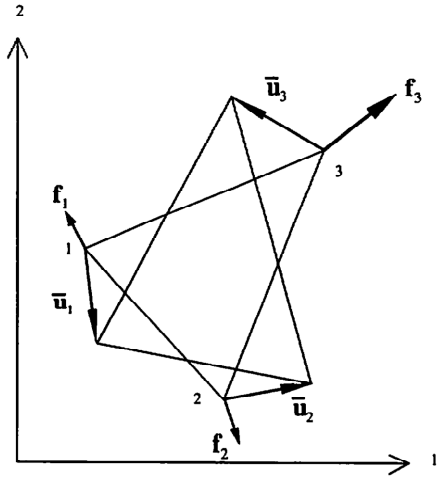



**Figure A1.** Force vector and virtual displacement vector work at each nodal point in a certain finite element (Hayashi, 2008).

Then, to obtain the practical form of (1), we assume the displacement within element as a function of coordinates. Since the simplest relation is linear, we take linear relation as follows.

$$\left(\bar{\mathbf{u}}^e\right)^T \mathbf{f}^e = \int_s \left(\bar{\mathbf{e}}^e\right)^T \mathbf{s} dS$$

Substituting the values of coordinate and displacement at nodes into this equation, we have

$$u_N = (1 \quad x_{N1} \quad x_{N2})\mathbf{a}$$

Writing in vector form

$$\mathbf{u}^e = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \end{bmatrix} \mathbf{a} = C\mathbf{a}$$

The coefficient vector **a** is derived from the equation,

$$\mathbf{a} = C^{-1}\mathbf{u}^e$$

$$C^{-1} = \frac{1}{\Delta}\begin{bmatrix} \Delta_{11} & \Delta_{21} & \Delta_{31} \\ \Delta_{12} & \Delta_{22} & \Delta_{32} \\ \Delta_{13} & \Delta_{23} & \Delta_{33} \end{bmatrix}$$

and $\Delta \equiv \det C$ and $\Delta_{ij} \equiv cofactor\ of\ C$.

Therefore, the inner displacement is represented in terms of nodal displacements

$$\begin{aligned} u &= \mathrm{c}\, C^{-1}\mathrm{u}^e \\ &= \frac{1}{\Delta}(\Delta_{11} + \Delta_{12}x_1 + \Delta_{13}x_2 \quad \Delta_{21} + \Delta_{22}x_1 + \Delta_{23}x_2 \\ &\quad \Delta_{31} + \Delta_{32}x_1 + \Delta_{32}x_2)\mathrm{u}^e \end{aligned}$$

Replacing as $\phi_N = \frac{1}{\Delta}(\Delta_{N1} + \Delta_{N2}x_1 + \Delta_{N3}x_2)$, we have $u = \phi_N u_N$.

Since we will consider 2D situation, displacement has 2 components as $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

$$u_1 = \phi_N u_{N1}$$

$$u_2 = \phi_N u_{N2}$$

Writing them in vector form,

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_1 & \phi_2 & \phi_3 \end{bmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$$

Then, exchanging the order of nodal displacements,

$$u_{11}u_{21}u_{31}u_{12}u_{22}u_{32} \Rightarrow u_{11}u_{12}u_{21}u_{22}u_{31}u_{32}$$

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{bmatrix} \phi_1 & 0 & \phi_2 & 0 & \phi_3 & 0 \\ 0 & \phi_1 & 0 & \phi_2 & 0 & \phi_3 \end{bmatrix} \begin{pmatrix} u_{11} \\ u_{12} \\ u_{21} \\ u_{22} \\ u_{31} \\ u_{32} \end{pmatrix}$$

$$= \Phi\mathbf{u}^e$$

Then, we can represent strain by nodal displacements as

$$\mathbf{e}=\begin{pmatrix}e_{11}\\ e_{22}\\ 2e_{12}\end{pmatrix}=\begin{pmatrix}u_{1,1}\\ u_{2,2}\\ u_{1,2}+u_{2,1}\end{pmatrix}=\begin{bmatrix}\phi_{1,1} & 0 & \phi_{2,1} & 0 & \phi_{3,1} & 0\\ 0 & \phi_{1,2} & 0 & \phi_{2,2} & 0 & \phi_{3,2}\\ \phi_{1,2} & \phi_{1,1} & \phi_{2,2} & \phi_{2,1} & \phi_{3,2} & \phi_{3,1}\end{bmatrix}\mathbf{u}^e$$

$$=B\mathbf{u}^e$$

Where

$$B=\frac{1}{\Delta}\begin{bmatrix}\Delta_{12} & 0 & \Delta_{22} & 0 & \Delta_{32} & 0\\ 0 & \Delta_{13} & 0 & \Delta_{23} & 0 & \Delta_{33}\\ \Delta_{13} & \Delta_{12} & \Delta_{23} & \Delta_{22} & \Delta_{33} & \Delta_{32}\end{bmatrix}$$

As for stress vector, according to the constitutive law of elasticity,

$$\mathbf{s}=\begin{pmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{12}\end{pmatrix}=D\mathbf{e}$$

For example, in case of plane strain

$$D=\frac{E(1-\nu)}{(1+\nu)(1-2\nu)}\begin{bmatrix}1 & \frac{\nu}{1-\nu} & 0\\ \frac{\nu}{1-\nu} & 1 & 0\\ 0 & 0 & \frac{1-2\nu}{2(1-\nu)}\end{bmatrix}$$

Then, according to the principle of virtual work,

$$=\left(\bar{\mathbf{e}}^e\right)^T\left[\int_S B^T\ DB\ dS\right]\mathbf{u}^e$$

$$\mathbf{f}^e=K^e\mathbf{u}^e$$

This is called the stiffness equation of element.

Superposing every stiffness equations of element, we obtain the stiffness equation of whole domain. **F** = **K u**

$$\left(\bar{\mathbf{u}}^e\right)^T\left(\mathbf{f}^e-\int_S\mathbf{f}_b dS\right)=\int_S\left(\bar{\mathbf{e}}\right)^T\mathbf{s}\,dS$$

If body force (fb) is considered, the principle of virtual work need be modified as

## 2. Fault Analysis

As shown in **Figure A2**, the Mohr-Coulomb criterion is written as a linear relationship between shear and normal stresses,

$$\sigma^*=\nu(\sigma_1+\sigma_2) \tag{2}$$

When we consider the analysis in plane strain condition, it is possible to calculate the value of third principal stress ($\sigma^*$), where $\nu$ is the Poisson ratio (Timoshenko and Goodier, 1970). After comparing the values of $\sigma_1$, $\sigma_2$ and $\sigma^*$, we can recognize the newly defined $\sigma_1$, $\sigma_2$ and $\sigma_3$ as the maximum, intermediate and mini mum principal stresses respectively. We introduce how

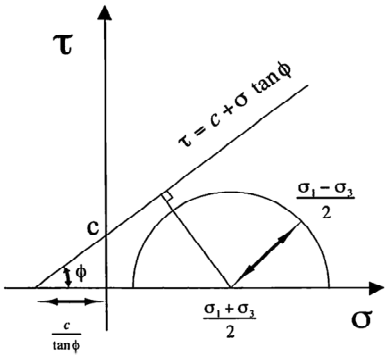



**Figure A2.** Failure envelope and Mohr's circle in σ-τ space. *c* is cohesion and φ is angle of internal friction (Hayashi, 2008).

mum principal stresses respectively. We introduce how the Mohr-Coulomb criterion is combined into the FE software package; though I already wrote the method of failure analysis in my serial papers (**Table 1**).

If body force ($\mathbf{f}_b$) is considered, the principle of virtual work need be modified as

$$\tau=c+\sigma\tan\phi \tag{3}$$

where $c$ and $\phi$ are the cohesive strength and the angle of internal friction, respectively. Failure will observe when the Mohr's circle first touches the failure envelope (3). It will happen when the radius of the Mohr's circle, $\sigma_1+\sigma_2/2$, is equal to the perpendicular distance from the center of the circle at $\sigma_1-\sigma_2/2$ to the failure envelope,

$$\left(\frac{\sigma_1-\sigma_3}{2}\right)_{failure}=c\cos\phi+\left(\frac{\sigma_1+\sigma_3}{2}\right)\sin\phi \tag{4}$$

According to Melosh and Williams (1989), the proximity to failure ($P_f$) is the ratio between the calculated stress and the failure stress, which is given by

$$P_f=\left[\frac{\left(\frac{\sigma_1-\sigma_3}{2}\right)}{\left(\frac{\sigma_1-\sigma_3}{2}\right)_{failure}}\right] \tag{5}$$

When the ratio reaches one ($P_f$ = 1), failure occurs, but when $P_f$ < 1 stress is within the failure envelope, rock does not fail. The proximity to failure $P_f$ reveals which parts of the model are close to failure or already failed by generating faults.

The type of faulting has been determined by the Anderson's theory (1951). According to his theory three classes of faults (normal, strike slip and thrust) result from the three principal classes of inequality that may exist between the principal stresses. I realized the judgment in the program *failure.state.func* in FE package.

# Call for Papers


# Natural Science

**A Journal Published by Scientific Research Publishing, USA**
**www.scirp.org/journal/ns**



Natural Science is an international journal dedicated to the latest advancement of natural sciences. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of natural sciences. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy. The journal publishes original papers including but not limited to the following fields:

- **Astronomy & Space Sciences**
  - Astronomy
  - Astrophysics
  - Atmospheric Science
  - Space Physics
- **Earth Science**
  - Geography
  - Geology
  - Geophysics/Geochemistry
  - Oceanography
- **Chemistry**
  - Analytical Chemistry
  - Biochemistry
  - Computational Chemistry
  - Inorganic Chemistry
  - Organic Chemistry
  - Physical Chemistry
- **Life Science**
  - Cell Biology
  - Computational Biology
  - Genetics
  - Immunology
  - Medicine/Diseases
  - Microbiology
  - Molecular Biology
  - Neuroscience
  - Pharmacology/Toxicology
  - Physiology
  - Psychology
  - Virology
- **Physics**
  - Applied Physics
  - Atomic, Molecular, and Optical Physics
  - Biophysics
  - High Energy/Particle Physics
  - Material Science
  - Plasma Physics
- **Others**
  - Education
  - History of Science
  - Science and Innovations

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

## Notes for Intending Authors

Submitted papers should not be previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. For more details, please access the website.

## Website and E-Mail

http://www.scirp.org/journal/ns ns@scirp.org

*Natural Science*, 2010, 2, *527*-666
Published Online June 2010 in SciRes (http://www.scirp.org/journal/ns).

# TABLE OF CONTENTS

**Volume 2 Number 6** **June 2010**

## LIFE SCIENCE



## PHYSICS



## OTHERS

