BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 10(6)

Comparative Estimation of Genetic Diversity in Population Studies using Molecular Sampling and Traditional Sampling Methods

Amr TM Saeb[1, 2]* & Satish Kumar David[2]

[1]Department of Entomology, Ohio State University, USA; [2]Information Technology Department, Strategic Center for Diabetes Research, King Saud University, Saudi Arabia; Amr TM Saeb – Email: saeb.1@osu.edu; Phone: +966114724179, Ext- 3334; Fax: +966114725682; *Corresponding author

Received June 03, 2014; Revised May 09, 2014; Accepted June 10, 2014; Published June 30, 2014

Abstract:
Entomopathogenic nematodes (EPN) are efficient biological pest control agents. Population genetics studies on EPN are seldom known. Therefore, it is of interest to evaluate the significance of molecular sampling method (MSM) for accuracy, time needed, and cost effectiveness over traditional sampling method (TSM). The study was conducted at the Mohican Hills golf course at the state of Ohio where the EPN *H. bacteriophora* has been monitored for 18 years. The nematode population occupies an area of approximately 3700 m^2 with density range from 0.25-2 per gram soil. Genetic diversity of EPN was studied by molecular sampling method (MSM) and traditional sampling method (TSM) using the mitochondrial gene pcox1. The MSM picked 88% in compared to TSM with only 30% of sequenced cox 1 gene. All studied genetic polymorphism measures (sequence and haplotype) showed high levels of genetic diversity of MSM over TSM. MSM minimizes the chance of mitochondrial genes amplification from non target organisms (insect or other contaminating microorganisms). Moreover, it allows the sampling of more individuals with a reliable and credible representative sample size. Thus, we show that MSM supersedes TSM in labour intensity, time consumption and requirement of no special experience and efficiency.

Keywords: Entomopathogenic nematodes, molecular population genetics, Genetic markers, ***pcox1***, molecular sampling, Genetic analysis, Bioinformatics analysis.

Background:
Entomopathogenic nematodes (EPN) and their symbiotic bacteria possess the ability to kill insect pests and therefore are efficient biological pest control agents **[1].** Together, they provide an emerging model system for research of important biological processes including pathogenesis, symbiosis and parasitism. However, genetic markers for entomopathogenic nematodes are limited and the tools of their population genetic studies are scarce. There are different genetic markers that can be utilized in molecular population genetics. For example, microsatellite markers sets that have been described for entomopathogenic nematodes *Heterorhabditis bacteriophora* **[2].** In addition, mitochondrial DNA (mtDNA) is widely used in population genetics investigations, phylogeny reconstruction and molecular evolution **[3, 4].** This is because mtDNA is highly variable, maternally inherited, lacks recombination and seems to be selectively neutral **[3, 5].** The elevated evolution rates of mtDNA genes permit their use in comparison of both inter- and intra-specific variation. Among mitochondrial genes, NADH dehydrogenase subunit 4 (*nd4*) and cytochrome c oxidase subunit 1 (*cox1*) provide ideal markers for population genetic structure and molecular evolution **[6, 7].** On the other hand, the complex life cycle of *Heterorhabditis* species has consequences on their natural populations, such as genetic

structuring, gene flow and genetic diversity. The alternation between sexual and asexual (hermaphroditic) reproduction in EPN should lead to continuous change of allele frequency and population structure **[8-10].** The is no doubt that population genetics studies in EPN is very complicated due to the microscopic size of the individuals, lack of morphological distinguishing characters, lack of suitable genetic markers, the need for host dissection to collect individuals and difficulty of purification of good quality DNA from individual nematodes. To address these problems that hinder EPN population genetics studies, we are suggesting a molecular sampling method for studying population genetics of EPN and comparing its resulting genetic diversity with traditional dissection method. Particularly, in the present study we used both morphological and molecular approaches to confirm identity of individual nematodes to the species level. Then, in order to investigate the genetic diversity between the molecular sampling method (MSM) and traditional sampling method (TSM), the *pcox1* genes was isolated, sequenced and analyzed.

Methodology:

The nematode population occupies an area approximately 3700 m² with a nematode density range from 0.25-2 per gram soil.

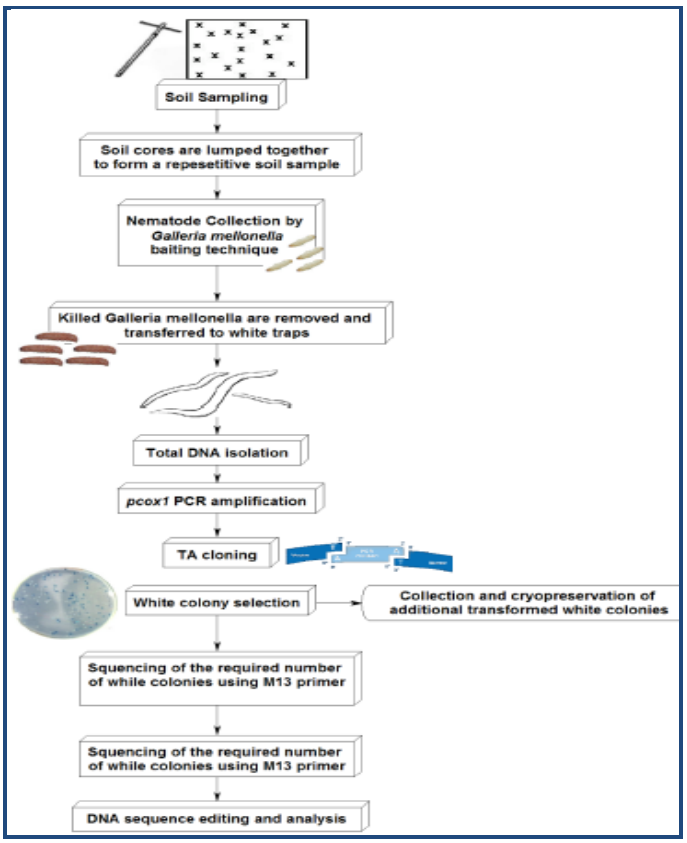

Figure 1: A summary of the suggested molecular sampling method.

Nematode Sampling

A total number of 15 sampling satiations were used to cover the Mohican Hills golf course at the state of Ohio, USA. A distance of 40 and 100 meters separated the sampling points. Nematodes were recovered from the soil samples using the larvae of wax moth *Galleria mellonella* baiting technique **[11].** Emerging nematodes were collected, washed three times with sterile distilled water and frozen at -80°C immediately untill used in the MSM (**Figure 1**). For comparison using the TSM, a total number of 50 individuals of first generation hermaphroditic nematodes were collected by dissection from the infected larvae and subjected to the molecular genetics analysis following Blouin *et al*. **[12].**

Molecular Genetics Analysis

For molecular sampling method, the nematode pellet was crushed into powder in liquid nitrogen and DNA was extracted using Qiagen® Genomic Tip 100/G according to the manufacturer's recommendations. Four µL of the supernatant was used to amplify the selected genes. For the traditional dissection method, 50 adults *H. bacteriophora* nematodes were collected from insect cadaver in placed in 25 µL lysis buffer **[2]** and incubated at –80 °C for 2 h, followed by incubation at 60 °C for 1 h and 95 °C for 15 min. For the specie confirmation, the ITS2 sequences of each individual nematode in TSM and collective MSM sample were amplified, sequenced and analyzed. The following primer set were used Forward: 5′-ACGTCTGGTTCAGGGTTGTT-3′and Reverse: 5′-TTAGTTTCTTTTCCTCCGCT-3′. For the amplification of the *pcox1* gene primers: Forward: 5′-TTT TTT GGG CAT CCT GAG GTT TAT-3′ and Reverse: 5′-AAA GAA AGA ACA TAA TGA AAA TG-3′ were used. PCR amplification products were cleaned up using Qiagen® PCR purification kit. The PCR products were TA cloned then sequenced using M13 primer.

DNA Sequencing and Data Analysis

All DNA raw sequence data were edited to remove low quality sequences, using Bioedit sequence alignment editor **[13]** and subjected to BLAST (Basic Local Alignment Search Tool) to perform sequence similarity searches. Sequences were initially aligned using the Bioedit built-in clustal W program. Additional sequence alignments were performed using MAFFT multiple sequence alignment program **[14]** and with PRRN (the best-first iterative refinement strategy with tree-dependent partitioning) multiple sequence alignment **[15].** Resulting alignments were compared and the final alignments were improved manually and prepared in required formats. In order to select the best substitution model, we used FindModel program **[16].** For *cox1* we used Jukes-Cantor (AIC1 = 1209.691724 lnL = -604.845862) as the substitution model for *cox1*. Pairwise distance calculation was performed using Jukes-Cantor models and gaps, marked as (-), and missing data and stop codons marked as (?), were taken into account using the complete deletion method in MEGA software **[17]** and the default gap values in PAUP V4B10, MrBayes v3.1 and TreeBuzzel **[18**, **19].** Amino acid translations of partial nucleotide cox1 sequences were obtained and analyzed by MEGA 6 software. Poisson correction was used as amino acid substitution method. DNA sequence polymorphism analysis was performed using DnaSP 5.10.1 software **[20].** Over all pairwise comparisons of sequence differences (D) was determined using the formula D = 1–(M/L), where M is the number of alignment positions at which the two sequences have a base in common, and L is the total number of alignment positions over which the two sequences are compared. The computer software DnaSP 5.10.1 was used to calculate the haplotype diversity (H) using the formula: $H = n (1- \sum \chi i2) / (n-1)$, where χi is the frequency of the i-th haplotype and n is the sample size for tested sampling method.

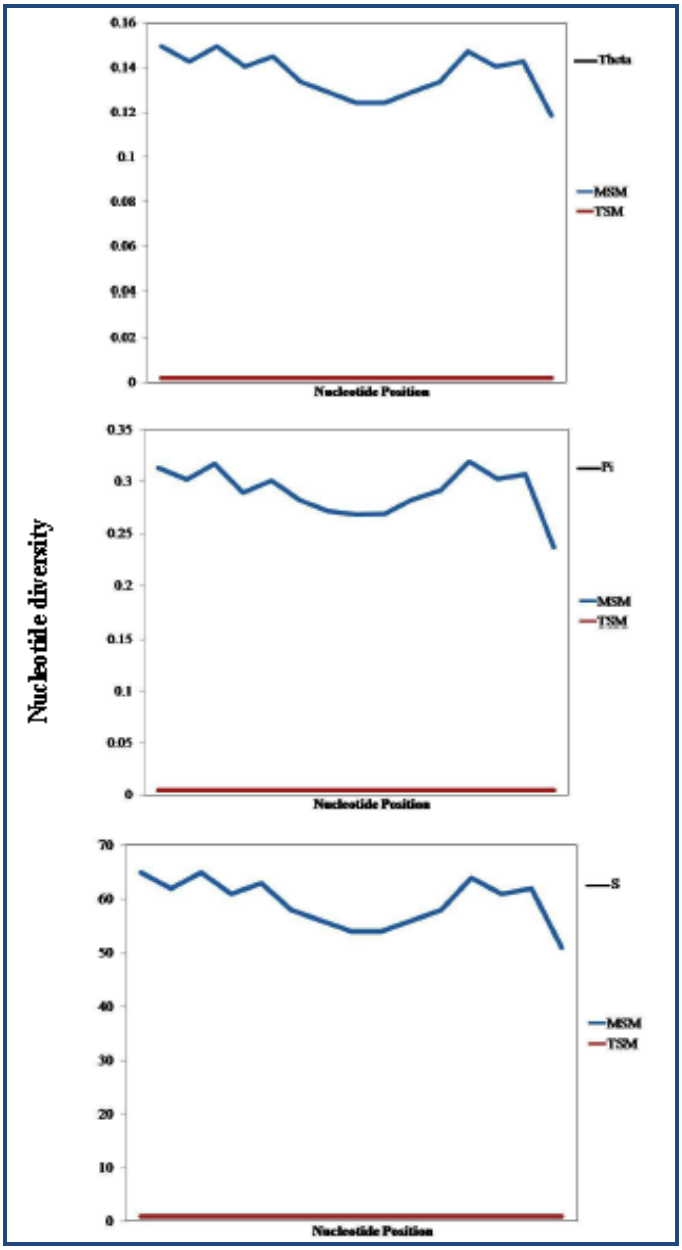

Figure 2: DNA Nucleotide diversity models, Pi (the average number of nucleotide differences per site between two sequences). Theta (calculated nucleotide diversity per site in the sequences) and S (the number of segregating sites in the sequences) for both molecular sampling method and traditional sampling method.

Results & Discussion:

Successful amplification of the target (pcox1) gene Molecular Sampling Method (MSM)

A total number of 50 *E. coli* positive white colonies were selected and processed as stated in the materials and method section. Of those, 44 high quality *cox1* nematode gene sequences were obtained. One insect *cox1* gene sequence, 3 amoeba and 2 non-recombinant plasmid (no insret = false positive) sequences.

Traditional Sampling Method (Dissection for Individual Nematodes) (TSM)

A total number of 50 individual first generation hermaphroditic nematodes were collected from the infected larvae and subjected to the molecular genetics analysis following Blouin *et al.* (1999) **[12].** Of those 50 individuals, only 15 high quality nematode sequences were obtained. 20 low quality sequences were observed, 5 insect *cox1* sequences, of lepidopteran origin, 7 *pcox1* of bacterial origin and 3 amoeba sequences.

Our results showed that, MSM success rate in obtaining the nematode *cox1* gene was 88%. Whereas, the TSM was only successful in obtaining the nematode *cox1* gene in 30% of the sequenced amplicons. Thus, suggested MSM is more efficient in acquiring the nematode target gene compared. This also implies that MSM is more cost effective to be used in molecular population genetic studies compared with TSM. Furthermore, in TSM, the amplification of *cox1gene* from untargeted organisms such as the insect host, symbiotic bacteria and contaminating amoeba emphasize on the failure of the removal of contaminates of insect tissues and bacteria or amoeba cells using the traditional dissection method despite the several washing steps of the hermaphroditic nematodes. While, in case of the molecular sampling method, the white trapping and the collective washing steps were able to remove greater part of the non-target tissues and cells before DNA isolation step, leaving behind only nematode templates for *pcox1* gene PCR amplification. Moreover, dissection and follow-up cleaning steps to at least 5-8 minutes for each infected cadaver to obtain the 50 individuals required for the study.

Genetic Diversity Evaluation

Table 1 shows a comparative analysis between polymorphic sites achieved via MSM and TSM. Results in **Table 1 (see supplementary material)** showed that Molecular Sampling Method was able to provide higher levels of variability in DNA sequence levels compared to the traditional method. For instance, the total number of observed mutations in the gene using molecular sampling method was 288 while it was zero in case of traditional method. Similar results were obtained from the DNA polymorphism analysis and nucleotide diversity **Table 2 (see supplementary material) & Figure 2.** Higher genetic polymorphism was observed in the MSM compared to the TSM that exhibited no genetic polymorphism in *pcox1* in the tested individuals. For example, nucleotide diversity value (Pi) for MSM was 0.28330 while, it was 0 in case of TSM. Moreover, overall average distances among tested sequences of MSM using Kimura 2-parameter model was 0.4902 while it was 0 in TSM.

In addition, **Table 3 (see supplementary material)** shows a haplotypic diversity comparison between MSM and TSM. The number of observed genetic haplotypes among sequences obtained by MSM was 8 while only 1 haplotype was observed using TSM. After considering the sites with alignment gap number of observed genetic haplotypes among sequences obtained by MSM was 11 while only 2 haplotype was observed using TSM. In both cases the haplotypic diversity was higher in case of MSM compared to TMS. Moreover, **Table 4 (see supplementary material)** representing Genetic code, Genetic recombination analysis and amino acid comparisons between MSM and TSM. Similarly, results in table 4 further demonstrate the ability of MSM to produce and detect genetic variability among parasitic nematode compared to TSM. For instance, the minimum number of recombination events detected using MSM were three, namely, between sites **[11, 12],** (146,154) and

(154,167). On the other hand no recombination events have been detected using TSM. In addition, deduced amino acid sequence analysis showed that using MSM, we detected 111 Parsimony informative sites, while, have been detected using TSM. Thus, MSM superseded TSM in all levels of genetic diversity investigated in this study.

There is no doubt that our suggested MSM has many advantages over the TSM. For example, MSM is greatly less labor intensive, less time consuming and requires no specific experience with nematode identification using morphological traits which is very intricate in the case of *Heterorhabditis* EPN and many other parasitic nematode species which are extremely morphologically conserved. Moreover, MSM enable us to effortlessly seizing for maximum number of genetic make-ups in the sampling site by treating larger number of nematodes from single or multiple infected insect cadavers. Moreover, MSM allows identification of more haplotypes and DNA sequence variability by allowing processing more positive colonies containing different haplotypes. Furthermore, this method allows revisiting the –80ºC preserved white transformed colonies in case of technical difficulties and/or searching for more DNA variability. In addition, MSM allows precise identification of the nematode species targeting the ITS1 region, which is not possible in case of using individual nematode which dose provide enough material for both molecular identification and pcox1 analysis. Furthermore, this method eliminates majority of molecular contamination problems faced when using chelex based methods. Moreover, it allows sampling more individuals to produce a more reliable and credible representative sample. Last but not the least, MSM allows identification of the presence of cryptic species in our sample and eliminating it from our analysis.

Conclusion:
We showed that MSM supersedes TSM in labour intensity, time consumption and requirement of no special experience. Moreover, it is more efficient in exhibiting genetic diversity among parasitic nematode used in this study.

Abbreviations:
Entomopathogenic Nematodes (EPN), Molecular Sampling Method (MSM), Traditional Sampling Method (TSM), mitochondrial DNA (mtDNA), NADH dehydrogenase subunit 4 (*nd4*), cytochrome c oxidase subunit 1 (*cox1*).

Competing Interests:
The authors declared that they have no competing interests period.

Authors' Contributions:
ATMS, designed the study collected samples, performed the molecular analysis, performed phylogenetic and bioinformatics analysis and prepared the manuscript; **SKD,** assisted in collecting information, performing phylogenetic and bioinformatics analysis and preparing the manuscript.

Acknowledgment:
We like to thank all members of the Information Technology Department for facilitating the conduction of the data analysis. We also want to thank the Department of Entomology, Ohio State University for facilitating field work.

References:
[1] *Shapiro-Ilan D et al. J Nematol*. 2014 **46:** 27 [PMID: 24643501]
[2] Bai X *et al. Mol Ecol Resour*. 2009 **9:** 207 [PMID: 21564604]
[3] Liu J *et al. J Parasito*. 1999 **85:** 709 [PMID: 10461953]
[4] Hu M *et al. Electrophoresis*. 2002 **23:** 27 [PMID: 11824618]
[5] Mattiucci S *et al. J Parasitol*. 2014 **100**: 199 [PMID: 24224764]
[6] Blouin MS, *Int J Parasitolo* 2002 **32:** 527 [PMID: 11943225]
[7] Hawdon JM *et al. Mol Ecol*. 2001 **10:** 1433 [PMID: 11412366]
[8] Hebert PD *et al. Proc Natl Acad Sci USA***.** 2004 **101:** 14812 [PMID: 15465915]
[9] Hebert PD, *Experientia Suppl*. 1987 **55:** 175 [PMID: 2961596]
[10] Majeský L, PLoS One. 2012 **7**: e41868 [PMID: 22870257]
[11] Sturhan D & Mrácek Z, *Folia Parasitol (Praha)***.** 2000 **47:** 315 [PMID: 11151957]
[12] Blouin MS *et al. Heredity (Edinb)***.** 1999 **83:** 253 [PMID: 10504422]
[13] Hall TA. *Nucleic Acids Symposium Series* 1999 **48:** 95
[14] http://align.genome.jp/mafft/.
[15] http://align.genome.jp/prrn/.
[16] http://hcv.lanl.gov/content/hcv-db/findmodel/findmodel.html
[17] Tamura K *et al. Mol Biol Evol* 2013 **30:** 2725 [PMID: 24132122]
[18] Swofford D. L. 2002. PAUP* software, version 4b10. Sinauer Associates, Sunderland, MA.
[19] Strimmer K & haeseler A. *Mol Biol Evol*. 1996 **13:** 964
[20] Rozas J, *Methods Mol Biol* 2009 **537:** 337 [PMID: 19378153]

Edited by P Kangueane
Citation: **Saeb & David,** Bioinformation 10(6): 347-352 (2014)

BIOINFORMATION

open access

Supplementary material:

Table 1: Polymorphic sites statistics of cox1 gene using Molecular Sampling Method and Traditional Sampling Method

Character	Molecular Sampling Method	Traditional Sampling Method
Number of sites:	550	533
Invariable (monomorphic) sites:	185	514
Variable (polymorphic) sites:	264	0
Total number of mutations:	288	0
Singleton variable sites:	0	0
Parsimony informative sites:	264	0
Singleton variable sites (two variants):	0	0
Parsimony informative sites (two variants):	240	0
Singleton variable sites (three variants):	0	0
Parsimony informative sites (three variants):	24	0
Variable sites (four variants):	0	0

Table 2: DNA sequence polymorphism comparison of *cox1 gene* between Molecular Sampling method and Traditional sampling method.

Character	Molecular Sampling Method	Traditional Sampling Method
Number of polymorphic (segregating) sites, **S:**	264	0
Total number of mutations, **Eta:**	288	0
Standard Deviation of Haplotype diversity:	0.04	0
Nucleotide diversity, Pi:	0.28	0
Theta (per site) from Eta:	0.15	0
Theta (per site) from S, Theta-W:	0.14	
Variance of theta (no recombination):	0.001	0
Standard deviation of theta (no recombination):	0.04	0
Variance of theta (free recombination):	0.0001	0
Standard deviation of theta (free recombination):	0.01	0
Finite Sites Model		
Theta (per site) from **Pi**:	0.45	0
Theta (per site) from **S**:	0.20	0
Theta (per site) from **Eta:**	0.18	0
Average number of nucleotide differences, **k:**	127	0
Stochastic variance of k (no recombination), **Vst(k):**	2951	0
Sampling variance of k (no recombination), **Vs(k)**:	142	0
Total variance of k (no recombination), **V(k):**	3093	0
Stochastic variance of k (free recombination), **Vst(k)**:	42	0
Sampling variance of k (free recombination), Vs(k):	2	0
Total variance of k (free recombination), V(k):	44	0
Theta (per sequence) from S, **Theta-W**:	61	0
Variance of theta (no recombination):	304	0
Variance of theta (free recombination):	14	0
Overall average distances: (Kimura 2-parameter):	0.5	0

Table 3: Haplotypic diversity comparison of *cox1 gene* between Molecular Sampling Method and Traditional Sampling Method.

Character	Molecular Sampling Method	Traditional Sampling Method
Number of Haplotypes, ***h*:**	**8** **Hap1**: 4 [1 3 5 25] **Hap2**: 15 [2 10-11 19-22 29-32 34 36 41 43] **Hap3**: 6 [4 6-9 35] **Hap4**: 4 [12-14 37] **Hap5**: 3 [15-16 38] **Hap6**: 3 [17-18 44] **Hap7**: 6 [23-24 33 39-40 42] **Hap8**: 3 [26-28]	**1** **Hap1**: 15 [1-15]
Haplotype (gene) diversity, ***Hd***	0.84	0
Variance of Haplotype diversity:	0.001	0

Number of Haplotypes, ***h***:	**11**	**2**
Sites with alignment gaps is considered	**Hap1**: 4 [1 3 5 25] **Hap2**: 9 [2 21-22 29-32 34 41] **Hap3**: 6 [4 6-9 35] **Hap4**: 3 [10-11 43] **Hap5**: 4 [12-14 37] **Hap6**: 3 [15-16 38] **Hap7**: 3 [17-18 44] **Hap8**: 3 [19-20 36] **Hap9**: 3 [23-24 39] **Hap10**: 3 [26-28] **Hap11**: 3 [33 40 42]	**Hap1**: 6 [1 3 9 11 13 15] **Hap2**: 9 [2 4-8 10 12 14]
Haplotype diversity, ***Hd***:	0.91	0.51

Table 4: Genetic code, recombination and amino acid polymorphism comparisons of cox1 gene between Molecular Sampling Method and Traditional Sampling Method.

Character	**Molecular Sampling Method**	**Traditional dissection method**
Total number of analyzed codons:	133	172.8
Effective number of codons, ENC:	46.7	60.8
Codon Bias Index, CBI:	0.55	0.5
G+C content at second codon positions, G+C2:	0.34	0.43
Average number of sites: 132.82:	132.82	168.80
G+C content at (synonymous) third codon positions, G+C3s:	0.23	0.36
Average number of sites:	125.5	165.80
G+C content at coding positions, G+Cc:	0.32	0.34
Average number of sites:	398.45	506.40
G+C content (selected region), G+C:	0.32	0.34
Average number of sites:	449	521.60
Minimum number of recombination events (Hudson and Kaplan 1985) **Rm:**	3	**0**
Recombination has been detected between sites:	(11,12) (146,154) (154,167)	**-**
Number of polymorphic (segregating) sites in table: **Substitutions considered: All Substitutions and Sites with alignment gaps were excluded**	264	0
Number of polymorphic (segregating) sites in table: **Substitutions considered: All Substitutions Sites with alignment gaps were included**	365	19
Number of polymorphic (segregating) sites in table: **Substitutions considered: All Substitutions Sites with alignment gaps were included if there is a polymorphism**	272	0
Amino acid analysis:		
Conserved sites, **C**:	71/183	170/177
Variable sites, **V**:	111/183	0
Parsimony informative, **PI**:	111/183	0
Singletons, **S**:	0	0