



HAL
open science

Caractères du génome mitochondrial

Cyril Gallut

► **To cite this version:**

Cyril Gallut. Caractères du génome mitochondrial. Biosystema, 18 (Caractères), pp.53-60, 2000.
hal-04041508

HAL Id: hal-04041508

<https://hal.science/hal-04041508v1>

Submitted on 17 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

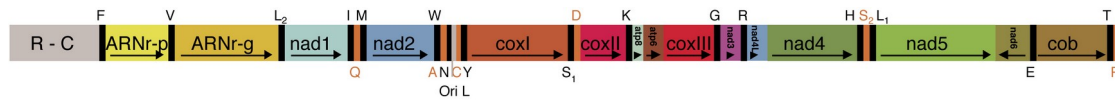
CARACTÈRES DU GÉNOME MITOCHONDRIAL

par Cyril GALLUT

Institut de Systématique (CNRS FR 1541)
CEB (JÉ 2160)
Laboratoire Informatique & Systématique
Université Pierre et Marie Curie
12 rue Cuvier 75005 Paris, France
gallut@ccr.jussieu.fr

Le génome mitochondrial des métazoaires est généralement constitué d'une unique molécule d'ADN circulaire dont la taille varie de 13 kb (KEDDIE *et al.*, 1998) à environ 40 kb (La ROCHE *et al.*, 1990). Son contenu en gènes est très conservé et se compose (figure 1) de treize gènes codant pour des sous-unités de protéines des complexes de l'oxydation phosphorylante, [cytochrome *c* oxydase sous-unités I-III (cox I-III), cytochrome *b* apoenzyme (cob), NADH déshydrogénase sous-unités 1-6 et 4L (nad1-6, 4L) et ATP synthase sous-unités 6 et 8 (atp6, atp8)], de deux sous-unités d'ARN ribosomique [petite et grande sous-unités (ARNr-p et ARNr-g)] et vingt-deux ARN de transfert utilisés dans la traduction des protéines codées par l'ADNmt. Généralement, le génome mitochondrial contient aussi une ou plusieurs régions non codantes dont une seule dépasse la dizaine de paires de bases, chez les vertébrés (CLAYTON, 1991) ainsi que chez certains arthropodes (GODDARD & WOLSTENHOLME, 1978). Il a été montré que cette région (région de contrôle) joue un rôle dans l'initiation de la réplication et de la transcription. Selon l'espèce, les gènes peuvent être portés par l'un des deux brins de la molécule uniquement ou bien

par les deux brins. Un très petit nombre d'espèces parmi celles étudiées à ce jour présentent des variations par rapport à ce schéma général. Certaines ont perdu un ou plusieurs de leurs gènes, d'autres ont un ADNmt linéaire constitué d'une ou deux molécules (BRIDGE *et al.*, 1992). Aujourd'hui le génome mitochondrial complet est connu chez plus de quatre-vingt-cinq espèces animales (confère BOORE, (1999) pour une revue récente), pour la plupart desquelles, l'ordre des gènes est rigoureusement le même. Un ensemble non monophylétique de vertébrés a conservé le même ordre (chondrichthyens, osteichthyens, mammifères etc.). L'embranchement le mieux représenté est celui des chordés, avec plus de la moitié des espèces. Ce biais d'échantillonnage s'explique principalement par le fait que le séquençage d'un génome mitochondrial n'est généralement pas motivé par la phylogénie mais par d'autres buts, comme l'étude d'un système génétique simple par exemple. Dans ce travail, un seul représentant par groupe d'espèces dont le génome est identique a été retenu, soient vingt-cinq espèces, représentatives de neuf phylums sur la trentaine connue.



Ceratotherium simum

Figure 1 : Génome mitochondrial du rhinocéros blanc *Ceratotherium simum*

cox I-III : cytochrome c oxydase sous-unités I-III, cob : cytochrome b apoenzyme, nad1-6, 4L : NADH déshydrogénase sous-unités 1-6 et 4L, atp6, atp8 : ATP synthase sous-unités 6 et 8, ARNr-p et ARNr-g : ARN ribosomique petite et grande sous-unités, R-C : région de contrôle, Ori-L origine de répliation du brin léger, les ARNt sont représentés par le code à une lettre de l'acide aminé auquel ils correspondent.

Pour exploiter les informations relatives à l'ordre des gènes du génome mitochondrial dans une perspective phylogénétique, différentes approches ont été envisagées, approches tant phénétiques que cladistiques. Les approches phénétiques utilisent des distances d'édition basées sur des opérateurs tels que la délétion, l'insertion, l'inversion, et la transposition (SANKOFF *et al.*, 1992) ou bien, sur les réarrangements des points de cassure (BLANCHETTE *et al.*, 1996 ; BLANCHETTE *et al.*, 1999 ; SANKOFF & BLANCHETTE, 1998a). Diverses approches cladistiques peuvent être envisagées telle que l'identification des synapomorphies à « l'œil » ou bien différentes méthodes de codage concernant les points de cassure, les réarrangements ou les rangs des gènes dans le génome. La première approche a été utilisée pour la phylogénie des arthropodes et des échinodermes (BOORE *et al.*, 1995 ; BOORE *et al.*, 1998 ; SMITH *et al.*, 1993), même s'il ne s'agit pas d'une approche cladistique à proprement parler. En effet il n'y a pas de mise en congruence des caractères par une analyse de l'ensemble des caractères décrits sur les taxons. Le codage des points de cassure a été testé par Sankoff (SANKOFF et BLANCHETTE, 1998b) pour le comparer avec les résultats donnés par sa propre méthode. Quant au codage des réarrangements, il pose un problème de raisonnement circulaire ; il est en effet nécessaire d'identifier les réarrangements entre les taxons étudiés, ce qui nécessite au préalable de connaître la phylogénie. Or il n'est pas admissible de se baser sur la phylogénie pour la reconstruire. De plus, il est difficile d'envisager tous les réarrangements possibles *a priori*. Mais surtout comment choisir parmi les divers scénarios possibles, celui que l'on va retenir pour l'analyse ? Le codage du rang des gènes dans le génome pose également des problèmes. D'une

part, l'hypothèse d'homologie entre le rang d'un même gène au sein de deux génomes est hasardeuse, d'autre part de façon pratique, on ne dispose pas d'une origine dans le génome mitochondrial qui soit partagée par l'ensemble des taxons et à partir de laquelle on puisse définir les rangs des différents gènes.

Je propose donc un nouveau codage de l'ordre des gènes qui repose sur leurs positions relatives, ce qui permet de tenir compte de la topologie des génomes pour reconstruire une phylogénie.

LE CODAGE

Afin de décrire au mieux l'organisation du génome mitochondrial, trois types de caractères distincts sont utilisés, des caractères de « présence-absence » du gène, des caractères de « sens » du gène et des caractères de « position relative » du gène au sein du génome.

Sur l'ensemble des métazoaires, le nombre et la nature des gènes présents dans le génome mitochondrial n'est pas tout à fait constant ; par exemple, *Mytilus edulis* et les nématodes, ont perdu l'atp8 (HOFFMANN *et al.*, 1992 ; OKIMOTO *et al.*, 1992). Pour tenir compte de l'absence de gènes chez certaines espèces, un caractère de type binaire, à l'état présent ou absent, est introduit.

Les gènes mitochondriaux peuvent être transcrits à partir d'un brin ou à partir du brin opposé ; ainsi pour chaque gène, un caractère binaire est également utilisé : 0 si le gène est transcrit à partir du brin principal, 1 dans le cas contraire (figure 2).

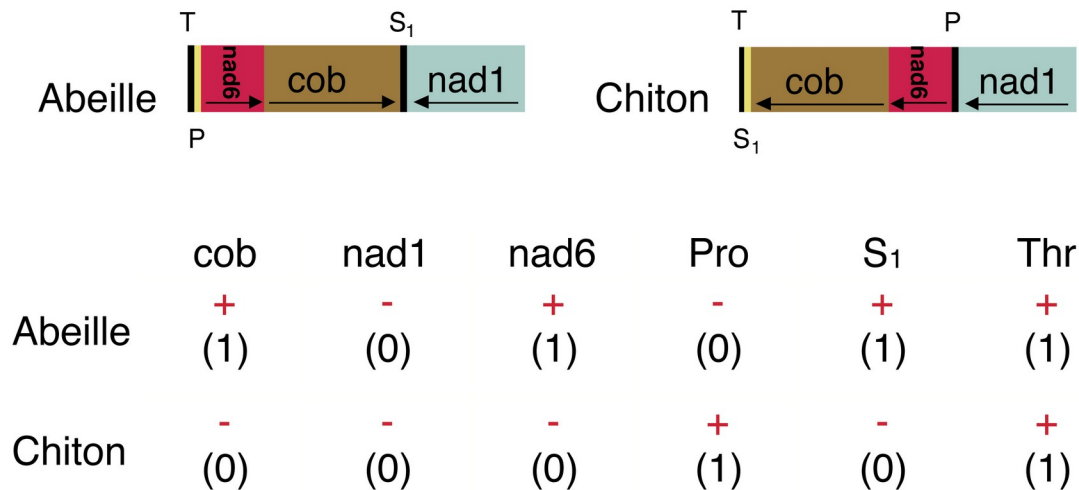


Figure 2 : Codage du sens du gène.

Pour décrire la position relative d'un gène dans la molécule, un caractère à états multiples est employé pour chaque gène. La position relative d'un gène pour une espèce donnée est décrite par ses deux voisins dans le génome (figure 3). Par conséquent, les états du caractère « position relative » d'un gène sont donnés par chacune des paires de gènes qui encadrent le gène dans les différents génomes étudiés. Les ARNt ont une mobilité plus importante que les gènes de grande taille, par conséquent la position relative de ces derniers sans tenir compte des ARNt renferme une information phylogénétique plus profonde. Pour tenir compte de cette information supplémentaire, un caractère de position sans les ARNt (codé de la même façon que les caractères de position relative) a été ajouté pour chaque gène de grande taille. Cela revient à coder chaque génome après avoir éliminé les ARNt.

Lorsqu'un gène est absent chez un taxon, j'introduis un point d'interrogation pour l'état de ce gène dans le codage des caractères de sens et de position, de cette manière, l'absence n'est prise en compte qu'une seule fois dans le

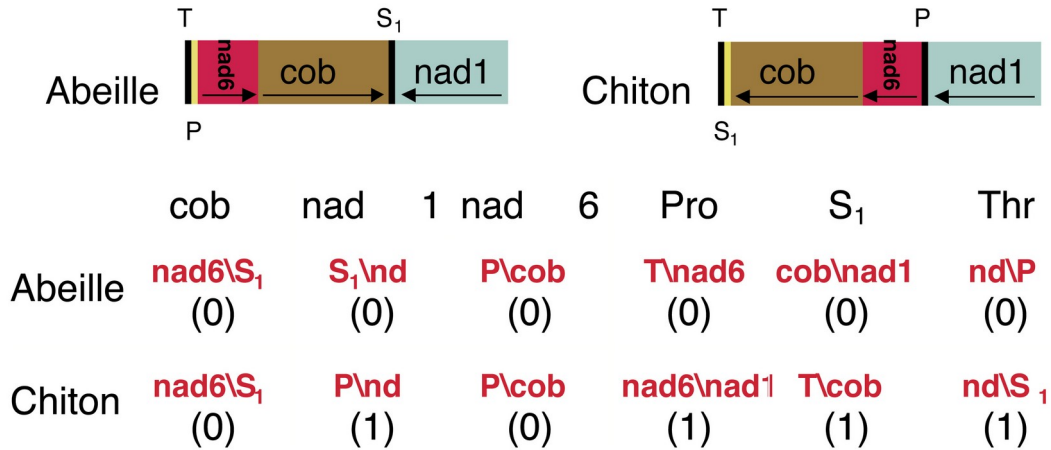
codage et l'on ne rajoute aucun pas supplémentaire (BARRIEL et TASSY, 1993).

En définitive, la matrice est constituée de vingt-cinq taxons et quatre-vingt-quinze caractères.

- Trois caractères de présence-absence (pour atp8, la région de contrôle et pour l'origine de réplication du brin léger).
- Trente-sept caractères pour le sens du gène.
- Trente-neuf caractères pour la position relative (la région de contrôle et l'origine de réplication n'ont pas de sens de transcription).
- Seize caractères de position relative sans les ARNt.

Cette méthode de codage permet de décrire le génome de chaque espèce indépendamment dans la matrice, sans connaissance de leurs relations phylogénétiques *a priori*. La méthode

de codage proposée ici permet de se placer dans un cadre cladistique, où les réarrangements peuvent être reconstitués *a posteriori* sur le cladogramme obtenu.



Position relative = les deux voisins

Figure 3 : Codage de la position du gène. nd : non disponible.

Quelles sont les implications de ce codage en termes de coût évolutif (nombre de pas) ? Il implique pour l'inversion, la transposition et la délétion un coût qui varie en fonction du nombre de gènes impliqués dans le réarrangement (figure 4). Par exemple l'inversion d'un gène ne coûte qu'un pas. En effet, il n'y a que le caractère de sens qui change d'état. Par contre, pour l'inversion de deux gènes contigus, non seulement les deux caractères de sens de ces gènes changent d'état, mais il y a aussi quatre caractères de position qui changent d'état (ceux des deux gènes inversés ainsi que ceux des gènes qui les encadrent). La transposition de trois gènes contigus coûte 6 pas. Six gènes différents changent de position relative : les deux gènes qui encadraient le bloc de trois gènes avant sa transposition, les deux gènes qui

encadrent le bloc après et enfin les deux gènes qui se trouvent aux extrémités du bloc. Le ou les gènes qui se trouvent à l'intérieur du bloc transposé ne changent pas de position relative puisqu'ils sont toujours encadrés par les mêmes gènes. Ceci explique pourquoi le nombre de pas ne change pas pour la transposition d'un bloc de deux gènes ou plus. Enfin, l'insertion ou la délétion d'un bloc coûte 2 pas plus 1 pas pour chaque gène du bloc, puisque les gènes qui encadrent le bloc inséré ou perdu changent de position relative et que les caractères de présence-absence de chacun des gènes changent d'état. Ces coûts ne sont pas représentatifs des modalités d'évolution du génome mitochondrial, dont on ne sait que peu de chose, mais ils ne sont que la conséquence du codage employé.

| Nombre de gènes par bloc | Inversion | Transposition | Délétion |
|--------------------------|--|---------------|--------------|
| 1 | 1 pas Sens : 1 pas | 5 pas | 3 pas |
| 2 | 6 pas Position : 4 pas Sens : 2 pas | 6 pas | 4 pas |
| 3 | 7 pas Position : 4 pas Sens : 3 pas | 6 pas | 5 pas |
| 4 | 8 pas Position : 4 pas Sens : 4 pas | 6 pas | 6 pas |

Figure 4 : Coût des différents réarrangements en fonction du nombre de gènes par bloc.

RÉSULTATS

La figure 5 montre le consensus strict des 36 arbres les plus parcimonieux de 611 pas (IC = 0,90 IR = 0,88), obtenus par branch and bound (implémenté dans PAUP (SWOFFORD, 1993)) avec la matrice de 95 caractères et 25 taxons. Sur cet arbre, les deutérostomiens sont monophylétiques, les hémichordés sont groupe frère des chordés, les mollusques groupe frère des arthropodes et enfin les ecdysozoa (AGUINALDO *et al.*, 1997) et les protostomiens sont paraphylétiques.

Les résultats préliminaires (figure 6), obtenus en codant les génomes tels qu'ils sont disponibles dans les banques de séquences, ne permettaient pas de retrouver la monophylie des mollusques (monophylie pourtant bien établie sur la base de caractères morphologiques (BRUSCA & BRUSCA, 1990)). Ce résultat a soulevé le problème de l'homologie des brins de la molécule d'ADN mitochondrial. En effet,

la prise en compte du sens des gènes implique que les brins qui portent le même gène, transcrit dans le même sens chez deux espèces différentes, doivent être considérés comme homologues. Il n'est possible d'établir cette homologie sur la base de critères biologiques (présence de la région de contrôle) que chez les chordés et les arthropodes. Il s'est donc avéré nécessaire de tester les deux possibilités pour chacun des autres taxons, en considérant qu'un brin était homologue du brin principal des vertébrés dans le premier cas et homologue du brin opposé dans l'autre cas. Les trente-deux combinaisons ont été testées indépendamment, en ne retenant au final que la matrice pour laquelle les mollusques sont monophylétiques (figure 5). Ce résultat a permis de montrer que les deux brins d'ADNmt des gastéropodes et du bivalve avaient été identifiés à l'inverse par rapport aux autres taxons dans les travaux précédents.



Consensus strict des 36 arbres les plus parcimonieux de 611 pas
 IC = 0,90 IR = 0,88

Figure 5 : Phylogénie des métazoaires basée sur l'ordre des gènes mitochondriaux. « Groupe vertébrés » représente l'ensemble (non monophylétique) des espèces de vertébrés qui ont le même ordre de gènes.

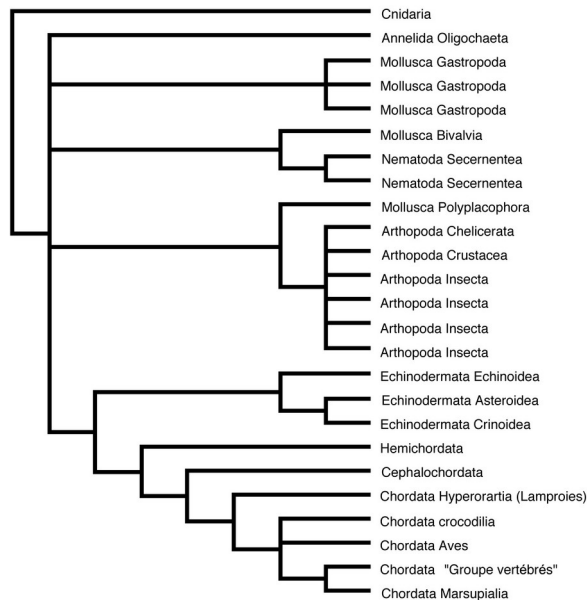


Figure 6 : Arbre préliminaire obtenu en codant les génomes tels qu'ils ont été publiés.

La reconstruction des réarrangements *a posteriori* permet de détecter des événements impossibles à identifier *a priori*. Dans l'exemple de la figure 7, on pourrait penser qu'il n'y a eu qu'un seul événement d'inversion pour passer du génome des arthropodes à celui du chiton. La reconstitution des événements sur l'arbre montre qu'il faut trois événements distincts pour expliquer ce réarrangement : une transposition de S1 et une inversion du bloc - P, ND6, cob - d'une part et une inversion de S1 d'autre part. Ces événements ayant eu lieu sur des branches différentes. À partir de ces reconstitutions, on peut établir le morphotype ancestral d'un taxon, c'est-à-dire la carte hypothétique du génome mitochondrial d'un ancêtre commun.

CONCLUSION

Les résultats obtenus ont permis de soulever les problèmes d'homologie des brins du génome mitochondrial. Ainsi en absence de critère biologique pour identifier les brins, il est possible de se baser sur un critère phylogénétique. Par ailleurs, cette analyse confirme l'intérêt du génome mitochondrial pour l'étude de la phylogénie profonde des métazoaires. L'intérêt de ce type de données est d'autant plus crucial que la reconstruction des relations de parenté entre embranchements de métazoaires rencontre de nombreux problèmes tels que la saturation des séquences moléculaires ou l'établissement parfois problématique des homologies primaires en morphologie. L'apparente simplicité de ces données ne doit néanmoins pas cacher la nécessité d'un traitement rigoureux. Ainsi le codage des positions relatives des gènes permet la description des taxons étudiés dans une perspective phylogénétique, sans connaissance *a priori* de leurs relations de parenté, tout en tenant compte de la topologie du génome.

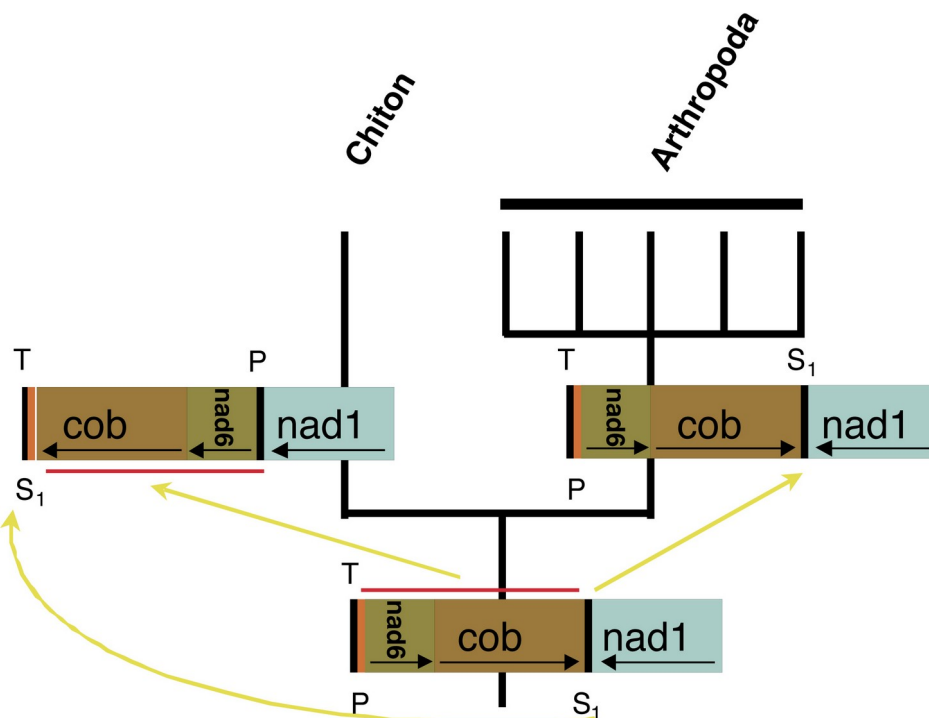


Figure 7 : Exemple de reconstitution des réarrangements sur l'arbre obtenu.

RÉFÉRENCES

- AGUINALDO, A. M., J. M. TURBEVILLE, L. S. LINFORD, M. C. RIVERA, J. R. GAREY, R. A. RAFF & J. A. LAKE. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387 : 489-493.
- BARRIEL, V. & P. TASSY. 1993. Characters, observations and steps: comment on Lipscomb's «parsimony, homology and the analysis of multistate characters». *Cladistics* 9 : 223-232.
- BLANCHETTE, M., T. KUNISAWA & D. SANKOFF. 1996. Parametric genome rearrangement. *Gene* 172 : GC 11-17.
- BLANCHETTE, M., T. KUNISAWA & D. SANKOFF. 1999. Gene Order Breakpoint Evidence in Animal Mitochondrial Phylogeny. *Journal of molecular evolution* 49 : 193-203.
- BOORE, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* 27 : 1767-1780.
- BOORE, J. L., T. M. COLLINS, D. STANTON, L. L. DAEHLER & W. M. BROWN. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376 : 163-165.
- BOORE, J. L., D. V. LAVROV & W. M. BROWN. 1998. Gene translocation links insects and crustaceans. *Nature* 392 : 667-668.
- BRIDGE, D., C. W. CUNNINGHAM, B. SCHIERWATER, R. DESALLE & L. W. BUSS. 1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proceedings of The National Academy of Sciences of The United States of America* 89 : 8750-8753.
- BRUSCA, R. C. & G. J. BRUSCA. 1990. Invertebrates. Sinauer associates, Sunderland Massachusetts. 922 pages.
- CLAYTON, D. A. 1991. Replication and transcription of vertebrate mitochondrial DNA. *Annu Rev Cell Biol* 7 : 453-478.
- GODDARD, J. M. & D. R. WOLSTENHOLME. 1978. Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. *Proceedings of The National Academy of Sciences of The United States of America* 75 : 3886-3890.
- HOFFMANN, R. J., J. L. BOORE & W. M. BROWN. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* 131 : 397-412.
- KEDDIE, E. M., T. HIGAZI & T. R. UNNASCH. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Molecular And Biochemical Parasitology* 95 : 111-127.
- LA ROCHE, J., M. SNYDER, D. I. COOK, K. FULLER & E. ZOUROS. 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. *Molecular Biology And Evolution* 7 : 45-64.
- OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY & D. R. WOLSTENHOLME. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130 : 471-498.
- SANKOFF, D. & M. BLANCHETTE. 1998a. Multiple genome rearrangement and breakpoint phylogeny. *Journal Of Computational Biology* 5 : 555-570.
- SANKOFF, D. & M. BLANCHETTE. 1998b. Comparative Genomics via Phylogenetic Invariants for Jukes-Cantor Semigroups In Proceedings of the International Conference on Stochastic Models.
- SANKOFF, D., G. LEDUC, N. ANTOINE, B. PAQUIN, B. F. LANG & R. CEDERGREN. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proceedings of The National Academy of Sciences of The United States of America* 89 : 6575-6579.
- SMITH, M. J., A. ARNDT, S. GORSKI & E. FAJBER. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *Journal of Molecular Evolution* 36 : 545-554.
- SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis Using Parsimony. Ver. 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign.