

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6575–6579, July 1992
Evolution

Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome

(genomics/algorithm/inversions/edit distance/conserved segments)

**DAVID SANKOFF*, GUILLAUME LEDUC*, NATALIE ANTOINE†, BRUNO PAQUIN†, B. FRANZ LANG†,
AND ROBERT CEDERGREN†**

**Presentation by Julie Hudson
MAT5313**

Definitions

- **Gene order:** permutation of genomic arrangement
- **Phylogenetics:** the study of evolutionary history and finding genetic connection between species
- **Mitochondrial genome:** complete set of genes specific to the mitochondria that guide its function

Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome

Research Question

Can we infer evolutionary history from the arrangement of genes in the mitochondria of various species?

Introduction

- Evolutionary inference traditionally done via comparison of homologous versions of a single gene
- mtDNA is susceptible to mutation (rapid nucleotide substitution) making homology difficult to differentiate from noise
- Genome level analysis will be more robust to this mutation
- Analytic tools different when comparing gene vs genome level similarity

Data

- Mitochondrial genome sequences from:

- Fungi (8)
- Animals (7)
- Protists (1)

With 31-50 mitochondrial genes (humans have 37)

- Why mitochondrial?

Complete genome consists of a small number of genes

Genome	Genes, no.
Fungi	
Fission yeast	
<i>Schizosaccharomyces pombe</i> (6)*	35
Budding yeast	
<i>Torulopsis glabrata</i> (7)	34
<i>Saccharomyces cerevisiae</i> (8)	39
<i>Kluyveromyces lactis</i> (9)	31
Filamentous Ascomycetes	
<i>Neurospora crassa</i> (10)	50
<i>Aspergillus nidulans</i> (11)	44
<i>Podospora anserina</i> (12)	45
Chytridiomycetes	
<i>Allomyces macrogynus</i> (B.P. and B.F.L., unpublished data)	35
Protist	
<i>Phytophthora infestans</i> (B.F.L., unpublished data)	31
Animals	
Vertebrates	
Mammalia (13, 14)	37
<i>Gallus gallus</i> (chicken) (15)	37
Echinoderms	
<i>Strongylocentrotus purpuratus</i> (sea urchin) (16)	37
<i>Asterina pectinifera</i> (star fish) (17)	37
<i>Pisaster ochraceus</i> (sea star) (18)	36
Insect	
<i>Drosophila yakuba</i> (19)	37
Nematode	
<i>Ascaris suum</i> (20)	36

Methods

- Edit Distance (noted as $E(a,b)$) is the number of elementary events necessary to change the gene order of one circular genome a into that of b .
- This is how gene arrangement is measured
- Elementary events: deletion, insertion, inversion, transposition.
- $E(a,b) = D(a,b) + R(a,b)$

Methods

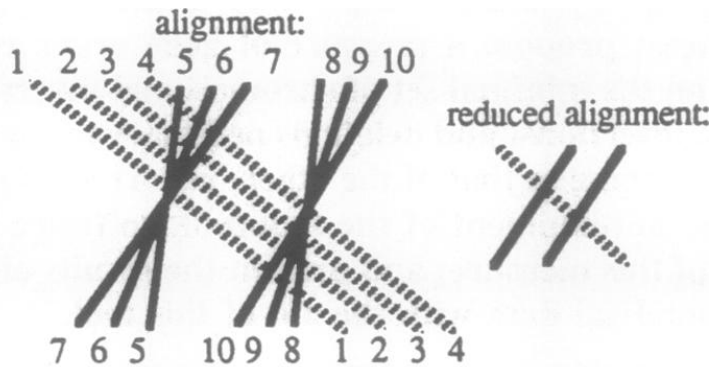
- $D(a,b)$ = total number of genes present in only one of a or b .
- (D for deletion/insertion)
- Simple to determine
- $a : 1, 2, 3, 4, 5 ; b : 1, 3, 4, 5, 6$
- $D(a,b) = 2$

Methods

- $R(a,b)$ = minimal number of inversion and transposition events necessary to convert one to the other, ignoring missing genes (D)
- Not so straight forward
- Based on a “conserved chromosomal segment” counting technique by Nadeau and Taylor, C
- $a : 1, 2, 3, 4, 5 ; \quad b : 1, 3, 2, 4, 5$
- $C = 3$ (what is expected from 1 inversion event)
- Differs if inversion occurs at the end or coincides with a previous inversion event.
- $a : 1, 2, 3, 4, 5 ; \quad b : 1, 2, 3, 5, 4$
- $C = 2$
- C is no greater than $2R$ in a circular genome

Methods

Alignment Reduction

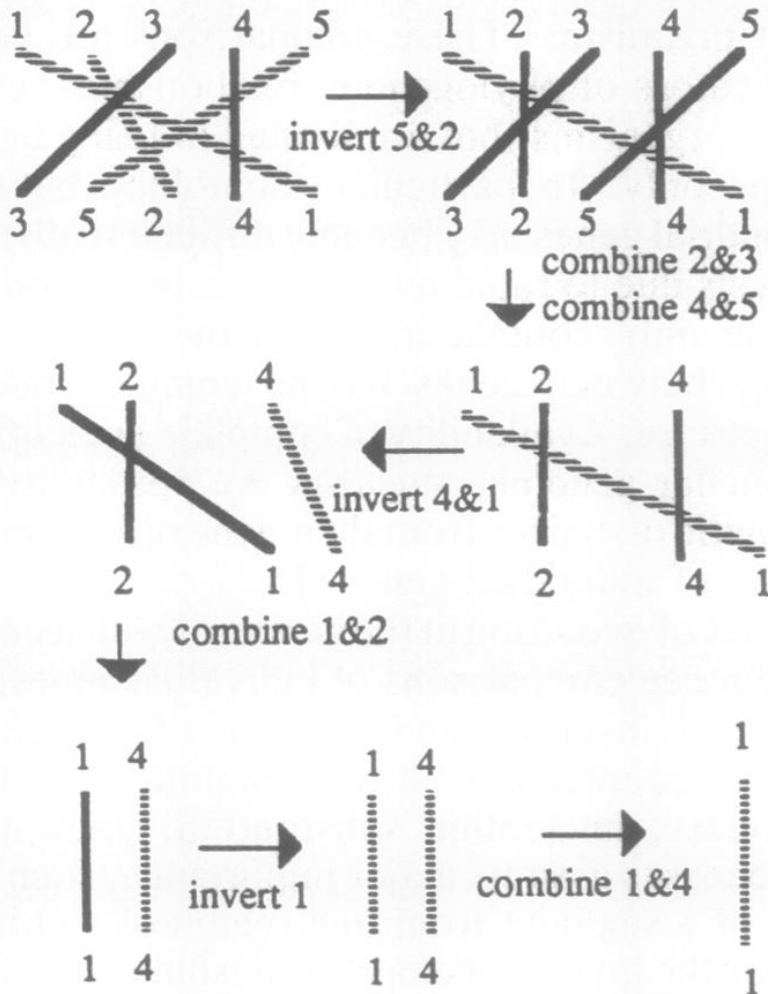


When pairs of genes are adjacent in both genomes with same orientation and order, they can be combined because the minimum number of recombinatory events will be the same. Conserved segments can be reduced

- Rearrangement Distance (R) determined through a branch-and-bound search using the program DERANGE.
- A series of alignment reductions and inversion/transposition events until the alignment is completely reduced (one link)

Methods

Three-inversion solution



DERANGE runs up to 5000 paths consecutively to determine which set of events returns the minimal rearrangement distance (where the bound-and-search is used).

Paths are discarded if they cannot lead to a minimal value or if the intermediate genome is determined to be probabilistically unlikely.




Example determining the minimal event distance through inversions and alignment reductions

Results

Distance (D below, R above)




	Mam	Gal	Str	Ast	Pis	Dro	Asc	Phy	All	Sch	Tor	Klu	Sac	Asp	Neu	Pod
Mam		1	18	16	19	13	25	12	18	21	17	16	19	23	26	27
Gal	0		19	17	17	12	26	13	22	21	17	17	19	23	24	26
Str	0	0		2	1	26	27	13	21	19	19	16	20	24	27	25
Ast	4	4	4		1	22	25	13	20	16	14	18	18	24	25	25
Pis	1	1	1	5		23	24	12	17	20	17	16	19	24	24	22
Dro	0	0	0	4	1		28	11	19	21	17	19	17	26	26	27
Asc	1	1	1	5	2	1		11	15	14	13	13	12	16	20	16
Phy	26	26	26	28	25	26	25		10	10	10	8	10	12	11	15
All	12	12	12	14	13	12	13	30		15	14	13	14	17	17	16
Sch	14	14	14	18	13	14	15	28	18		15	15	18	18	19	18
Tor	17	17	17	19	16	17	18	29	19	7		11	10	15	12	15
Klu	18	18	18	20	17	18	19	28	20	8	3		11	11	11	13
Sac	20	20	20	24	19	20	21	32	24	10	5	8		13	13	15
Asp	11	11	11	13	12	11	12	31	13	13	16	17	21		10	10
Neu	15	15	15	19	16	15	16	35	19	17	22	23	25	14		9
Pod	10	10	10	14	11	10	11	30	16	14	19	20	22	11	15	

Table 2 from paper: Distance between genome pairs

-  Animal
-  Protist
-  Fungus

Results




Edit Distance (D + R)																
	Mam	Gal	Str	Ast	Pis	Dro	Asc	Phy	All	Sch	Tor	Klu	Sac	Asp	Neu	Pod
Mam																
Gal	1															
Str	18	19														
Ast	20	21	6													
Pis	20	18	2	6												
Dro	13	12	26	26	24											
Asc	26	27	28	30	26	29										
Phy	38	39	39	41	37	37	36									
All	30	34	33	34	30	31	28	40								
Sch	35	35	33	34	30	35	29	38	33							
Tor	34	34	36	33	33	34	31	39	33	22						
Klu	34	35	34	38	33	37	32	36	33	23	14					
Sac	39	39	40	42	38	37	33	42	38	28	15	19				
Asp	24	34	35	37	36	37	28	43	30	31	31	28	34			
Neu	41	39	42	44	40	41	26	46	36	36	34	34	38	24		
Pod	37	36	36	39	33	37	27	45	32	32	34	33	37	21	24	

-  Animal
-  Protist
-  Fungus

Results

Edit Distance (D + R)

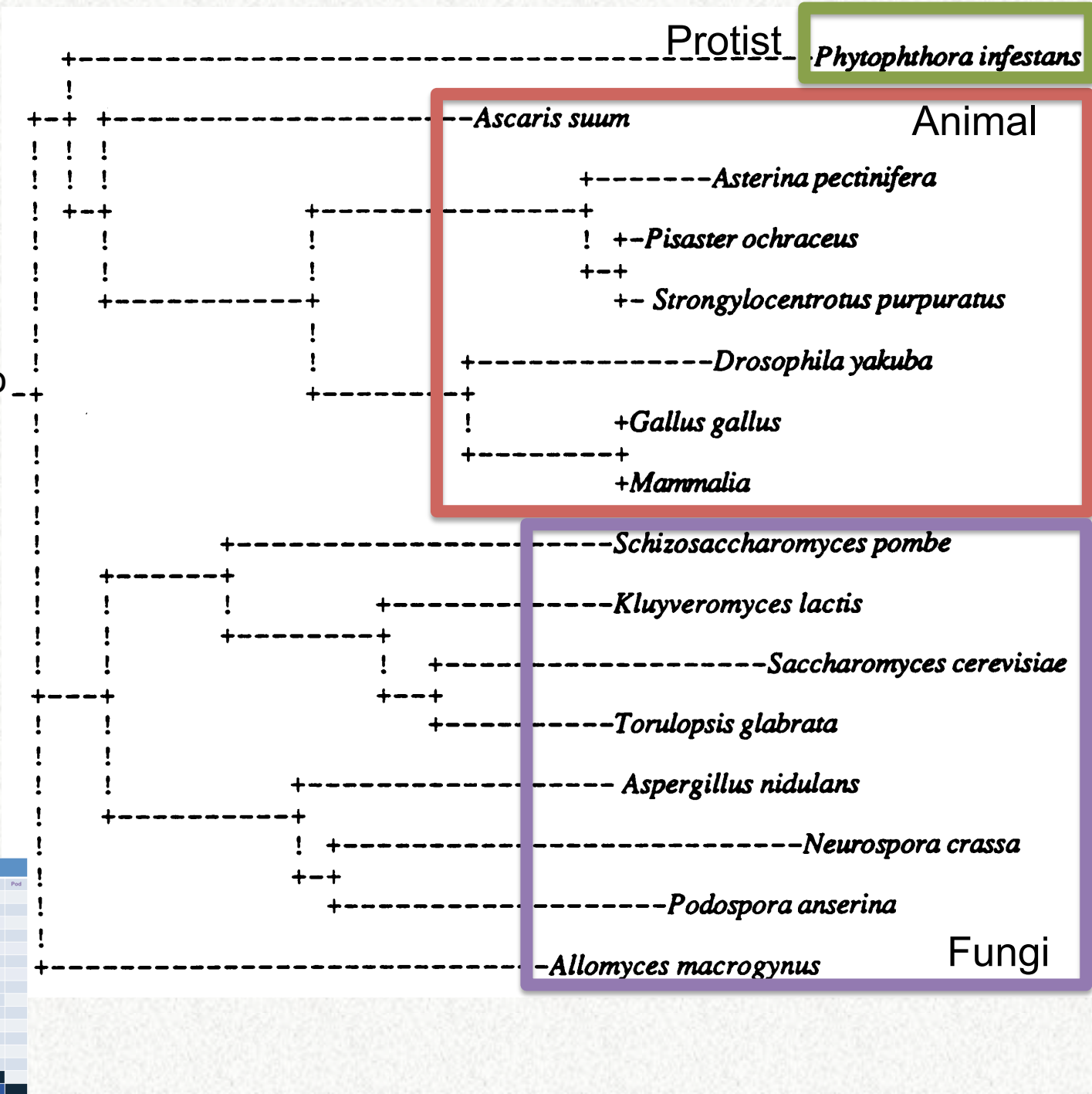
	Mam	Gal	Str	Ast	Pis	Dro	Asc	Phy	All	Sch	Tor	Klu	Sac	Asp	Neu	Pod
Mam																
Gal	1															
Str	18	19														
Ast	20	21	6													
Pis	20	18	2	6												
Dro	13	12	26	26	24											
Asc	26	27	28	30	26	29										
Phy	38	39	39	41	37	37	36									
All	30	34	33	34	30	31	28	40								
Sch	35	35	33	34	30	35	29	38	33							
Tor	34	34	36	33	33	34	31	39	33	22						
Klu	34	35	34	38	33	37	32	36	33	23	14					
Sac	39	39	40	42	38	37	33	42	38	28	15	19				
Asp	24	34	35	37	36	37	28	43	30	31	31	28	34			
Neu	41	39	42	44	40	41	26	46	36	36	34	34	38	24		
Pod	37	36	36	39	33	37	27	45	32	32	34	33	37	21	24	

-  Animal
-  Protist
-  Fungus

Dark to light -> Close to far

Results

Edit distances fitted to an additive tree model using a weighted least-squares criterion generated this tree



Results

- **Validation:** to show calculated $R(a,b)$ significantly different from random noise
- **Method:** random circular permutations created and tested like the genome to determine noise level
- **Result:** Within animal and fungi group the values of R are non-random; between the groups is random
- Shows that R contains phylogenetic information

Discussion

- **Overall assessment:** coherence of phylogeny indicates that the macrostructures of genomes contain quantitatively meaningful information for phylogenetic reconstruction
- **Assumptions:**
 - All inferred rearrangement events contribute the same amount to E (solution: weighting)
 - Each deletion is a separate event (solution: slow-growing convex function to determine D if events simultaneous)
 - No back mutation (solution: simulation study determining a correction for this underestimation)