doi:10.5598/imafungus.2018.09.01.04 IMA FUNGUS · 9(1): 37–48 (2018)

ARTICLE

Mitochondrial introgression and interspecies recombination in the *Fusarium fujikuroi* species complex

Gerda Fourie, Nicolaas A. Van der Merwe, Brenda D. Wingfield, Mesfin Bogale, Michael J. Wingfield, and Emma T. Steenkamp

Department of Genetics, Biochemistry and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa; corresponding author e-mail: gerda1.fourie@fabi.up.ac.za

Abstract: The *Fusarium fujikuroi* species complex (FFSC) is an economically important monophyletic lineage in the genus *Fusarium.* Incongruence observed among mitochondrial gene trees, as well as the multiple non-orthologous copies of the internal transcribed spacer region of the ribosomal RNA genes, suggests that the origin and history of this complex likely involved interspecies gene flow. Based on this hypothesis, the mitochondrial genomes of non-conspecific species should harbour signatures of introgression or introgressive hybridization. The aim of this study was therefore to search for recombination between the mitochondrial genomes of different species in the FFSC. Using methods based on mt genome sequence similarity, five significant recombinant regions in both gene and intergenic regions were detected. Using coalescent-based methods and the sequences for individual mt genes, various ancestral recombination events between different lineages of the FFSC were also detected. These findings suggest that interspecies gene flow and introgression are likely to have played key roles in the evolution of the FFSC at both ancient and more recent time scales.

Key words:
evolutionary history
FFSC
heteroplasmy-associated mitochondrial recombination
hybridization
introgression
species concepts

Article info: Submitted: 12 January 2018; Accepted: 18 February 2018; Published: 27 February 2018.

INTRODUCTION

The *Fusarium fujikuroi* species complex (FFSC, previously referred to as the *Gibberella fujikuroi* species complex) is one of several monophyletic assemblages in the genus *Fusarium* (phylum *Ascomycota*, order *Hypocreales*) (Geiser *et al.* 2013). This complex is well-known for the many well-documented plant pathogens and mycotoxin producers it includes (Kvas *et al.* 2009). Previous work suggests that the FFSC likely emerged during the middle-to-late Miocene (O'Donnell *et al.* 2013) and that its evolutionary history could have involved interspecies gene flow (O'Donnell & Cigelnik 1997). Such interspecies interactions have also been described from other *Fusarium* species (e.g. *F. oxysporum* and *F. graminearum* species complexes) (Ma *et al.* 2010, O'Donnell *et al*. 2000a) and, in the FFSC, was suggested to explain the existence of multiple non-orthologous copies of the internal transcribed spacer region of the ribosomal RNA genes (O'Donnell & Cigelnik 1997).

Interspecies gene flow is typically associated with hybridization and introgression (Stuckenbrock 2016). Hybridization is the production of viable and recombinant offspring by non-conspecific individuals. Introgression occurs when short-lived hybrids backcross with individuals from the parental species, allowing incorporation of new genetic material into the genome of that parental species. The process of introgressing new genetic material into the gene pool of a species is referred to as "introgressive hybridization" (Anderson & Hubricht 1938). In nature, interspecies gene flow is generally thought to be limited by species isolation mechanisms, such as vegetative incompatibility, pheromone-receptor recognition, intersterility and post-zygotic nuclear-cytoplasmic incompatibility systems that restrict or prevent the exchange of genetic material between species (Giraud *et al* 2008). In many *Fusarium* species, including those in the FFSC, laboratory-based mating studies have shown that the level of reproductive isolation is not complete and that various species are capable of interbreeding (Desjardins *et al.* 2000, Leslie *et al.* 2004b).

The mitochondrial (mt) genome is potentially a valuable tool for studying hybridization and introgression in fungi. Fungal mitochondria are mostly inherited from the maternal parent (Taylor 1986), but cross-species interactions would often lead to a short-lived heteroplasmic state in which the hybrid individual would harbour mt haplotypes from both parents (Ballard & Whitlock 2004, Barr *et al.* 2005). Recombination between the different haplotypes would cause the introduction and/or replacement of new genes/regions on one or both mt genomes. Such signatures of the ancestral cross-species interactions thus would be retained

in the species' mt genomes, despite the fact that one of the mt haplotypes would typically be purged from subsequent populations (Rand 2001). In fish, for example, Wilson & Bernatchez (1998) described an ancient introgession in *Salvelinus namaycush* (trout) due to the presence of a single mt haplotype belonging to *S. alpines* (arctic char) in the *S. namaycush* population. In plants, Jaramillo-Correa & Bousquet (2005) described mitochondrial recombination between *Picea mariana* (black spruce) and *P. rebens* (red spruce) as a result of introgressive hybridization in the zone of contact between these conifers in north-eastern North America. Examples from fungi are still limited, but Fourie *et al*. (2013) hypothesized that the incongruence observed among gene trees inferred from mitochondrial genes could have resulted from recombination between the mt genomes of non-conspecific species.

In this study we considered the hypothesis that introgression or introgressive hybridization occurred in the history of the FFSC. Our first aim was to identify and characterize regions in the mt genomes of extant FFSC species that potentially originate from such interspecies gene flow events. For this purpose, the mt genomes for two FFSC species (i.e. *F. mangiferae* and *F. sterilihyphosum*) were determined and used to complement those (*F. circinatum*, *F. verticillioides* and *F. fujikuroi*) already in the public domain (Al-Reedy *et al*. 2012, Fourie *et al*. 2013). These genomes were then subjected to the recombination detection method (Martin & Rybicki 2000), Bootscan (Martin *et al*. 2005), Geneconv (Padidam *et al*. 1999) and Maximum χ^2 (Smith 1992) analyses that were designed for detecting interspecific recombination (Martin *et al*. 2010). The second aim of this study was to utilize a coalescent-based approach for detecting ancestral recombination in the mt genes of extant FFSC species (Price & Carbone 2005). We purposefully did not employ phylogenetic methods given the low sequence diversity observed in the mt genes of the FFSC and other fungi (e.g. Seifert *et al*. 2007, Huang *et al*. 2008, Fourie *et al*. 2013). For these coalescent analyses, the sequences for five mt genes (*atp*6, *cox*2, *nad*3, *nad*5, and *nad*6), previously shown to support incongruent phylogenetic histories (Fourie *et al.* 2013), and from a collection of species spanning the diversity of the FFSC, were utilized. To assess the potential effects of false negatives and/or systematic errors (i.e. artefacts that arise from failure to fully account for the properties of these data) (Delsuc *et al.* 2005) in the analyses, the degree to which selection and substitution rate heterogeneity affected the individual mt gene datasets were also evaluated.

MATERIAL AND METHODS

Fungal isolates

Twenty-seven *Fusarium* isolates representing three, four and five species in the respective "African", "Asian" and "American" clades of the FFSC (O'Donnell *et al*. 1998, 2000b), were used (Table 1). This collection included the standard mating type tester strains for the nine mating populations (i.e. MP-A to MP-I) or biological species of the FFSC (Leslie & Summerell 2006, Kvas *et al.* 2009), as well as representatives of *F. mangiferae* and *F. sterilihyphosum.*

Mt genome sequencing and assembly

To determine the mt genome sequence for *F. temperatum* isolate CMWF 389 and *F. mangiferae* isolate CMWF 1214 (Table 1), total genomic DNA was extracted as described previously (Groenewald *et al*. 2006) and subjected to pyrosequencing at Inqaba Biotechnologies (Pretoria, South Africa) on a single lane using the GS-FLX platform (Roche 454 system, Life Sciences, CT). After exclusion of low quality reads, those encoding mt sequences were identified using BLAST comparison to the available FFSC mt genomes (Al-Reedy *et al*. 2012, Fourie *et al*. 2013). The mt reads for the two species were subsequently assembled *de novo* with the CLC Genomics Workbench software version 6.0 (CLC bio, Århus, Denmark). The order and orientation of contigs were determined using *F. circinatum*, *F. verticillioides,* and *F. fujikuroi* as reference genomes. Gaps between contigs were filled manually by Sanger sequencing. Protein coding and tRNA mt genes were identified with MFANNOT and RNAweasel (http://megasun.bch.umontreal.ca) (Lang *et al*. 2007), as well as tRNAscan-SE (Lowe & Eddy 1997). Gene identities were confirmed with BLASTp comparisons against NCBI.

Mt genome-based recombination analysis (RDP, Bootscan, Geneconv and Maximum χ^2)

The Recombination Detection Program (RDP) package version 3.44 (Martin *et al*. 2010) was used to screen for possible recombination events in the five FFSC mt genomes and the individual gene datasets using RDP (Martin & Rybicki 2000), Bootscan (Martin *et al*. 2005), Geneconv (Padidam *et al*. 1999) and Maximum χ^2 (Smith 1992). Since these tools differ with regards to their power to detect recombination (Wiuf *et al*. 2001, Posada 2002), results from all four recombination detection methods were compared and only recombination identified by all four methods were considered. For these analyses (see below), the five mt genomes and the individual gene datasets were aligned using the CLC Genomics Workbench software.

The RDP method identifies potential recombinant segments by plotting the pair-wise percentage identity values of all combinations of three sequences/isolates within the given dataset. A potential recombinant region is subsequently identified as the region where the pair-wise percentage identity of sequence A to C or B to C is higher than that of A to B given that A and B are more closely related to one another than to C. The probability that the potential recombinant occurred by chance is then approximated using the binomial distribution (Martin & Rybicki 2000). Bootscan identifies potential recombination segments by constructing pair-wise distances and bootstrap replicates within overlapping sequence blocks. High degrees of bootstrap support for different tree topologies suggest potential recombinant regions (Martin *et al*. 2005). Geneconv detects recombination by identifying aligned sequenced pairs, where a match between two sequences is given +1 and a mismatch is awarded a penalty -m. The mismatch penalty depends on the density (ratio) of polymorphic sites between the sequences and the mismatch intensity parameter (G-scale), which is proportional to the total number of site differences (i.e. polymorphic sites) between the two isolates (Padidam *et al*. 1999). The Maximum χ^2 test searches for recombination break points by comparing the number of segregating sites on both sides of a putative recombination break point and calculating

Table 1. Isolate information for the *Fusarium fujikuroi* complex species used in this study.

Species	MP [a]	Clade [b]	Isolate number [c]	Other number [d]	Accession number [e]	Reference
F. circinatum	H	American	CMWF 350	MRC 7870/Fsp34	JX910419*	Fourie *et al.* 2013
F. circinatum	H	American	CMWF 497	MRC 7488		This study
F. circinatum	H	American	CMWF 498	MRC 6213		This study
F. verticillioides	A	African		NRRL 29056	JN041210*	Al-Reedy *et al.* 2012
F. verticillioides	A	African	CMWF 1196	MRC 8559		This study
F. verticillioides	A	African	CMWF 1197	MRC 8560		This study
F. fujikuroi	C	Asian	CMWF 1220	IMI58289	JX910420*	Fourie *et al.* 2013
F. fujikuroi	C	Asian	CMWF 1200	MRC 8532		This study
F. fujikuroi	C	Asian	CMWF 1201	MRC 8534		This study
F. subglutinans	E	American	CMWF 1204	MRC 8553/6483		This study
F. subglutinans	E	American	CMWF 1205	MRC 8554/6512		This study
F. temperatum	E	American	CMWF 1206	MRC 1084		This study
F. temperatum	E	American	CMWF 389	MRC 7828	KP742837*	This study
F. mangiferae	N/A	Asian	CMWF 1213	MRC 8092/8093		This study
F. mangiferae	N/A	Asian	CMWF 1214	MRC 7559	KP742838*	This study
F. sacchari	B	Asian	CMWF 1198	MRC 8552		This study
F. sacchari	B	Asian	CMWF 1199	MRC 8551		This study
F. proliferatum	D	Asian	CMWF 1202	MRC 8549		This study
F. proliferatum	D	Asian	CMWF 1203	MRC 8550		This study
F. thapsinum	F	African	CMWF 1207	MRC 8558		This study
F. thapsinum	F	African	CMWF 1208	MRC 8557		This study
F. nygamai	G	African	CMWF 1209	MRC 8546		This study
F. nygamai	G	African	CMWF 1210	MRC 8547		This study
F. konzum	I	American	CMWF 1211	MRC 8545		This study
F. konzum	I	American	CMWF 1212	MRC 8544		This study
F. sterilihyphosum	N/A	American	CMWF 1215	MRC 2802		This study
F. sterilihyphosum	N/A	American	CMWF 1216	MRC 8105		This study

[a]Mating populations (i.e. MP-A to MP-I) or biological species of the FFSC (Leslie & Summerell 2006, Kvas *et al*. 2009).

[b]"African", "Asian" and "American" clade designation of the *Fusarium* species based on their associated plant hosts (O'Donnell *et al*. 1998).

[c]All isolates used in this study are maintained in the *Fusarium* Culture Collection of Mike Wingfield, FABI, University of Pretoria, South Africa.

[d]Additional culture collections: MRC = The South African Medical Research Council, NRRL = ARS Culture Collection, USDA, IMI = CABI Biosciences, Egham, UK.

[e]Accession numbers of the mitochondrial genome and gene sequences used in this study. Whole mt genome sequences are indicated with asterisks. Individual genes were deposited to the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/EMBL) under the accession numbers LN8111335-LN811269 in the gene order atp6, cox2, nad3, nad5 and nad6 and isolate order similar to Table 1.

2 x 2 χ^2 values as an expression of the difference on either side of the central partition (Smith 1992). The p-value was set to 0.05 for all methods employed.

Mt gene sequencing

Total genomic DNA was extracted from week-old cultures (Table 1) incubated on half strength potato dextrose agar (PDA; Biolab Diagnostics, Wadeville, South Africa) at 25 °C. For amplification of mt genes, Primer3 (http://primer3.sourceforge.net/) was used to design primers that target *nad*3, *nad*5 and *nad*6 encoding the respective nicotinamide adenine dinucleotide (NADH) dehydrogenase subunits, *apt*6 that encodes adenosine triphosphate (ATP) synthase subunit 6, and *cox*2 that encodes cytochrome c oxidase subunit II (Supplementary Table S1). PCR reaction mixtures were adjusted to 25µl with sterile distilled water and contained *ca.* 5 ng/ml DNA, 0.5 mM of each primer, 250 mM dNTPs (Fermentas, Nunningen, Switzerland), 0.04 U/ml *Taq* DNA polymerase (Roche Molecular Biochemicals, Manheim, Germany) and PCR buffer with $MgCl_2$ (Roche). PCR cycling conditions consisted of an initial denaturation step at 94 °C for 5 min, followed by 35 cycles at 94 °C for 35 s, each primer pair specific annealing temperature (Supplementary Table S1) for 35 s, and 72 °C for 90 s with a final extension step at 72 °C for 5 min. Amplification products were precipitated and purified with polyethylene glycol (Hartley & Bowen 2003) and sequenced in both directions using the original primers, the BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and anABI PRISM®377 DNA sequencer (Applied Biosystems).

Coalescent-based detection of recombination in mt genes

Individual sequence alignments were collapsed into binary matrices by excluding segregating sites and indels using SNAP MAP (Aylor *et al*. 2006). This was done in order to assume the infinite-sites model of mutation where at most one mutation event can occur at each site (Kimura 1969). The minimum number of recombination events (R_m) within each binary matrix (gene dataset) was determined using RECMIN (Myers & Griffiths 2002) in SNAP Workbench (Price & Carbone 2005). R_m is based on the four-gamete test of Hudson & Kaplan (1985) that infers recombination between pairs of loci at which all four possible gametic types are present. Finally, minimal ancestral recombination graphs (ARG) were reconstructed using the BEAGLE branch and bound algorithm (Lyngsø *et al*. 2005) in SNAP Workbench.

The sequence data for whole mt genomes were not used in these coalescent-based analyses. This is because the high sequence diversity of the intergenic and/or intron regions (Al-Reedy *et al*. 2012, Fourie *et al*. 2013) would increase the false positive recombination events detected under the infinite-site model of mutation (McVean *et al*. 2002). Conversely, the individual mt gene datasets were not subjected to the analytical tools included in RDP3. This is because of these tools have limited value for detecting recombination in highly conserved regions (Posada 2002, Tsaousis *et al*. 2005) such as the five mt gene datasets examined here (see below).

Evaluating possible sources of systematic error and/or false positives

The ability to detect recombination in DNA sequences depends on the genetic diversity of the data as well as among site rate variation (Posada 2002). Little genetic diversity within the dataset could obscure the signal for recombination whereas rate variation could allow recombination to be detected incorrectly. Nucleotide diversity, sequence divergence and rate heterogeneity of each of the individual mt gene datasets were, therefore, estimated. For each dataset, DNAsp ver. 5 (Librado & Rozas 2009) was used to determine π, which is the average number of nucleotide differences per site between two sequences (Nei 1987). This software package was also used to determine the sequence divergence estimates D_{xy} and D_a, which respectively are the average and net numbers of nucleotide substitutions per site between species (Nei 1987). D_{xy} and D_a were used to estimate divergence between species within the FFSC where *F. oxysporum* was used as the outgroup (Cunnington 2007, Pantou *et al*. 2008). For comparative purposes, π, D_{xy} and $D_a$values were converted to percentages. jModeltest was used to evaluate the pattern of among-site rate heterogeneity for all gene datasets by estimating the shape parameter (α) of the gamma distribution, where smaller α values indicate strong rate variation (Yang 1996, Posada 2008).

Signals of recombination might also be obscured by other evolutionary phenomena such as directional selection acting on the target genes and/or analytical artefacts arising from factors such as substitution saturation. Substitution saturation results in homoplasy (Rubinoff & Holland 2005), which can incorrectly point towards recombination because recurrent mutations (i.e. mutation hot spots) and recombination can generate similar patterns of genetic variability (Eyre-Walker *et al*. 1999, Hagelberg 2003, Galtier *et al*. 2006). In addition, recurrent mutation could also result from selection pressure acting on the target genes or selection pressures acting on specific regions of the target genes (Nielsen 2005, Reed &Tishkoff 2006). We, therefore, tested if positive selection acted on the mt gene datasets and determined the level of substitution saturation in the mt gene datasets.

Specific sites under positive or negative selection were identified using three codon based maximum likelihood methods. These included the Fixed Effect Likelihood (FEL), Random Effect Likelihood (REL) and Single Likelihood Ancestor Counting (SLAC) methods from Datamonkey (http://www.datamonkey.org/) developed by Kosakovsky *et al.* (2005). FEL estimates ω for each site in a sequence alignment. REL allows rate variation in both non-synonymous and synonymous rates and a general underlying nucleotide substitution model. SLAC reconstructs ancestral sequences using the joint likelihood reconstruction method in the codon-state space (Kosakovsky *et al.* 2005). Results arising from all methods were compared and only codons identified as being under selection by all methods were considered.

The level of substitution saturation was measured by calculating the information entropy-based index of substitution saturation (Xia *et al*. 2003) with DAMBE5 (Xia 2013). This is a tree-based approach where substitution saturation (I_{SS}) can be determined by testing if the observed entropy at site *i* is significantly smaller than the expected entropy under full substitution saturation. We compared I_{SS} to the critical I_{SS} value, $I_{SS.C}$, where the latter depends on the topology of the tree, the number of taxa, the sequence length, the nucleotide frequency, and the transition/transversion ratio, all of which are studied and compared through simulations of an experimental set of topologies given the alignment (Xia *et al*. 2003, Xia & Lemey 2009). Since the third codon is more variable due to the wobble effect of the genetic code (Spencer & Barral 2012) and thus likely to experience more substitutions, substitution saturation was determined for the first and second codons separately from the third codon of each gene dataset.

RESULTS

Mt genome sequencing and assembly

Pyrosequencing together with Sanger sequencing allowed for the assembly of the mt genomes of *Fusarium temperatum* and *F. mangiferae*. In both cases, the sequences spanned the entire replicon, except for a gap containing the large subunit ribosomal RNA gene and three of the clusters of tRNA genes, i.e. tRNA gene clusters 2, 3 and 4 (Fourie *et al*. 2013). None of the pyrosequencing reads mapped to the corresponding region in the mt genomes of *F. circinatum, F. fujikuroi,* and *F verticillioides*, in which this sequence has been determined. Also, the repetitive nature of these regions (Fourie *et al*. 2013) precluded their amplification and sequencing, despite various attempts using multiple primerpairs.

The mt genome sequences of *F. temperatum* (GenBank KP 742837) and *F. mangiferae* (GenBank KP 742838) contained the

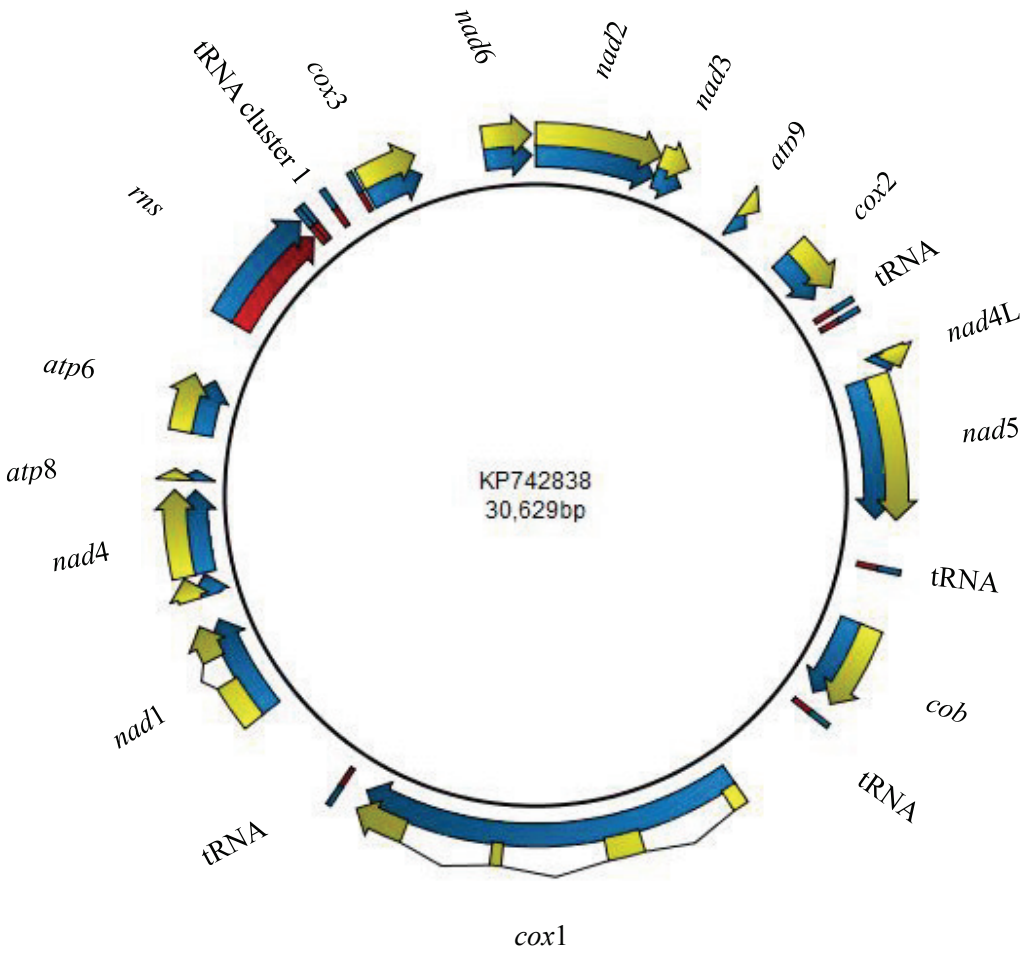


Fig. 1. Annotated map for the mitochondrial fragment of *Fusarium temperatum* (KP742837). The genome fragments encode the 14 protein coding genes of the oxidative phosphorylation pathway (blue = entire gene; yellow = coding sequence), one rRNA (red), 5tRNA (red) and tRNA cluster 1.

14 known mt protein coding genes (Fig. 1), the products of which are involved in oxidative phosphorylation. Within these protein coding genes, 12 and four group 1 introns were, respectively, found in the two mt genomes. *Fusarium temperatum* and *F. mangiferae* both contained an intron in their *cob* gene, while the *F. temperatum cox*1 gene contained 8 introns and the *cox*3, *nad*1 and *nad*2 genes each contained one intron, as opposed to the three introns in *cox*1of *F. mangiferae* as well as *cox*3 and *nad*2 that were free of introns (Table 2).

Although the presence of introns within protein coding genes varied greatly among and within the species examined, comparison of the five mt genomes suggested a possible link between intron abundance and the FFSC clade of the species (Table 2). For example, the mt genomes of the "American" clade species *F. circinatum* and *F. temperatum* both contained 14 introns as opposed to the one and four introns found in the mt genes of *F. fujikuroi* and *F. mangiferae,* respectively, that reside in the "Asian" clade and the three introns in the mt genes of the "African" clade species *F. verticillioides*. Regarding tRNA genes outside clusters 2, 3, and 4 (fragment not sequenced), the mt genome sequences of both *F. temperatum* and *F. mangiferae* contained tRNA cluster 1 which encodes four tRNA genes, as well as nine individual tRNA genes (Fig. 1). In both assemblies, all protein coding and sequenced tRNA genes were located in the same gene order and orientation (Fig. 1), similar to what has been described for other FFSC mt genomes (Al-Reedy *et al*. 2012, Fourie *et al*. 2013).

Mt genome-based recombination analysis

All four of the recombination detection tools identified recombinant regions within the FFSC genomes examined

Table 2. The number of introns identified in the mitochondrial protein coding genes of the *Fusarium* species used in this study.

	nad2	***cob***	***cox1***	***nad1***	***cox3***
F. circinatum	1	4	7	1	1
F. temperatum	1	1	8	1	1
F. fujikuroi	0	1	0	0	0
F. mangiferae	0	0	3	1	0
F. verticillioides	0	0	2	1	0

Table 3. The recombinant regions detected by RDP, Geneconv, Bootscan and Maximum χ^2 of the *Fusarium* mitochondrial genomes.

Recombinant[a]	Recombinant region[b]			Parents[c]	P-values [d]			
	Begin	End	Genes included		RDP	GeneConv	Bootscan	Maxχ^2
F. circinatum	20 366	20934	intergenic between tRNA.cys and *cox*1	*F. mangiferae* x unknown	6.89E-04	3.96E-02	4.54E-03	6.25E-05
F. mangiferae	2862	4530	*atp*9 to *cox*2	*F. fujkuroi* x *F. circinatum*	2.50E-45	2.29E-48	2.46E-46	8.77E-19
F. mangiferae	10621	10820	intergenic between *cob* and tRNA.cys	*F. fujikuroi* x unknown	1.03E-13	2.59E-11	8.52E-13	4.14E-05
F. mangiferae	23296	24471	*atp*6	*F. fujikuroi* x *F. circinatum*	3.90E-11	5.08E-08	2.29E-11	1.16E-06
F. temperatum	11143	11771	tRNA.arg	*F. mangiferae* x unknown	1.78E-22	2.54E-24	9.66E-27	2.47E-11

[a]The FFC species or daughter in which the recombinant region was detected.

[b]The position of the recombinant region according to the mt genome of each specific FFSC species, *F. mangiferae* (KP742837), *F. temperatum* (KP742838), and *F. circinatum* (JX910419).

[c]The suggested origin of the recombinant region according to RDP, Geneconv, Bootscan and Maximum χ^2 implemented in RDP version 3.44 (Martin *et al*. 2010). Unknown = the parent of this recombination event is not present within the dataset.

[d]The P-values for each of the recombination detection methods RDP, Geneconv, Bootscan and Maximum χ^2 implemented in RDP version 3.44 (Martin *et al*. 2010). Only those events detected with all four methods are shown (see Supplementary Table S3 for the results for all individual tests). RDP identifies recombinant segments via pair-wise percentage identity values (Martin & Rybicki 2000), Geneconv detects recombinant segments as aligned pairs that are unusually long and sufficiently similar (Padidam *et al*. 1999), Bootscan identifies recombinant segments as high degree of bootstrap support for different phylogenies (Martin *et al*. 2005) and Maximum χ^2 detects recombinant sections by comparing the number of segregating sites on both sides of the recombinant breakpoint (Smith 1992).

in this study (Supplementary Table S2). The consensus of the four detection methods suggested five significant recombinant regions in both gene and intergenic regions (Table 3). For example, recombinant regions were detected within the intergenic region between the tRNA gene for cysteine and the *cox*1 gene, as well as between *cob* and the tRNA gene for arginine. Recombination was also detected within the *atp*6, *atp*9 and *cox*2 genes. For all five of the detected recombination events, RDP, Bootscan, Geneconv, and Maximum χ^2 suggested events in which *F. circinatum*, *F. mangiferae* or *F. temperatum* were predicted to be the daughter of the recombination event, although the major and minor parents could not always be identified (Table 3).

Coalescent-based detection of recombination in mt genes

Within the five mt gene sequence datasets examined for the 27 *Fusarium* isolates included in this study, 17 recombination events were detected using RECMIN. The estimated minimum number of recombination events needed to explain the incompatibilities in the individual datasets were 0, 0, 1, 3, 3, and 10 for the *nad*3, *cox*2, *atp6, nad*6 and *nad*5 datasets, respectively. This suggested that recombination occurred in all the datasets examined with the exception of *nad*3 and *cox*2, and that recombination was extensive in *nad*5.

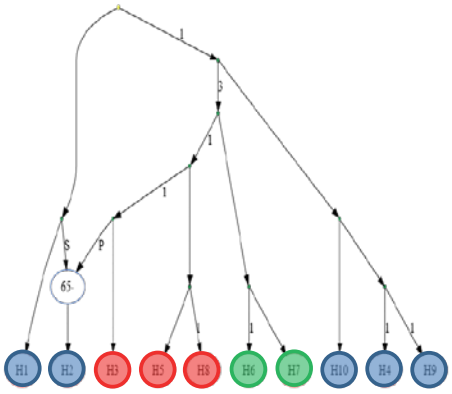
SNAP MAP collapsed the 27 sequences for each mt dataset into their respective haplotypes. There were six haplotypes in each of the *nad*3 and *cox*2 datasets, 10 in each of the *nad*6 and *atp*6 datasets and 15 in the *nad*5 dataset (Supplementary Table S3). Each haplotype typically comprised of the representatives for each species, or the representatives of closely related/sister species combined into a single haplotype (Supplementary Table S3).

Consistent with that suggested by the minimum number of recombination events, the ARG analysis suggested extensive recombination in the sequences of *nad*5, with some recombination in the sequences of *atp*6 and *nad*6 and no recombination in the sequences of *nad*3 and *cox*2 (Fig. 2). The ARG analysis also allowed identification of the recombinant region within each mt gene dataset. For example, it identified a recombination region within the *atp*6 dataset at nucleotide position 65 (Fig. 2, Supplementary Table S4).

The ARGs were also used to determine the relative order in which recombination occurred and to evaluate the contribution of mutations and coalescent events. Overall, the "American" clade haplotypes were associated with the deepest recombination events in the ARGs inferred from *nad*5 (position 1629 and position 1026) and *nad*6 (position 332) datasets. In addition, the ARGs of *atp6, nad*6 and *nad*5 suggested that recombination events had also occurred more recently in the evolutionary history of the FFSC based on their emergence towards the tips of the ARGs, for example, recombinant position 65 of the *atp*6 dataset and/or positions 503 and 515 of *nad*6 dataset. Finally, the ARGs of *nad*6 and *nad*5 also suggested that recombination occurred between the clades of the FFSC. For example, recombination event position 1629 of the *nad*5 dataset (haplotype H4) resulted

Fig. 2. Minimum ancestral recombination graphs (ARGs) inferred from the FFSC datasets for the *atp*6, *nad*5 and *nad*6 using the BEAGLE branch and bound algorithm in SNAP Workbench 2.0 (Price & Carbone 2005). Haplotypes are colour coded according to the clade within the FFSC complex to which the species belong ("African" = red, "American" = blue, "Asian" = green). Circles with numbers represent recombination events and the number within a circle represents the nucleotide position of the recombination event for each dataset. After a recombination event, the two sequences are replaced with a recombinant consisting of a prefix (P) from one sequence and a suffix (S) of the other sequence. The numbers on the branches suggest the number of mutation events before the coalescence of the specific haplotypes.

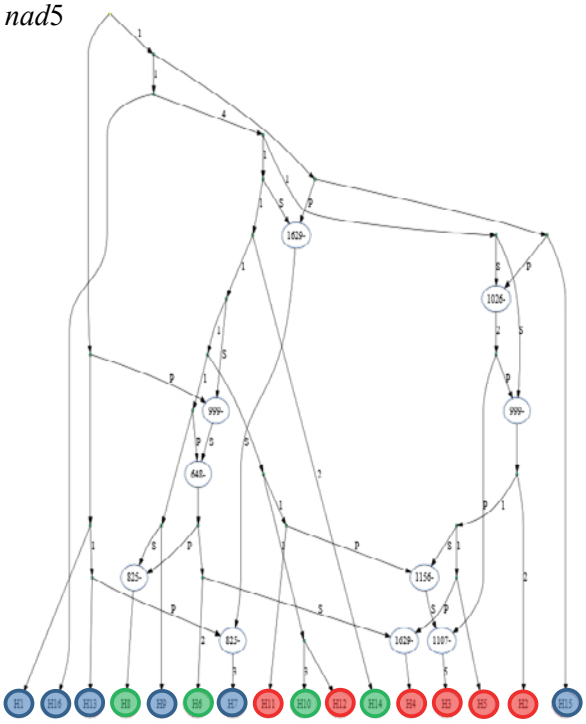
atp6



H1 = *F. circinatum*
H2 = *F. circinatum*
H3 = *F. nygamai*
H4 = *F. strelihyphosum*
H5 = *F. verticillioides*
H6 = *F. fujikuroi*
H7 = *F. mangiferae, F. sacchari, F. proliferatum*
H8 = *F. thapsinum*
H9 = *F. subglutinans, F. temperatum*
H10 = *F. konzum*

= American Clade
= African Clade
= Asian Clade

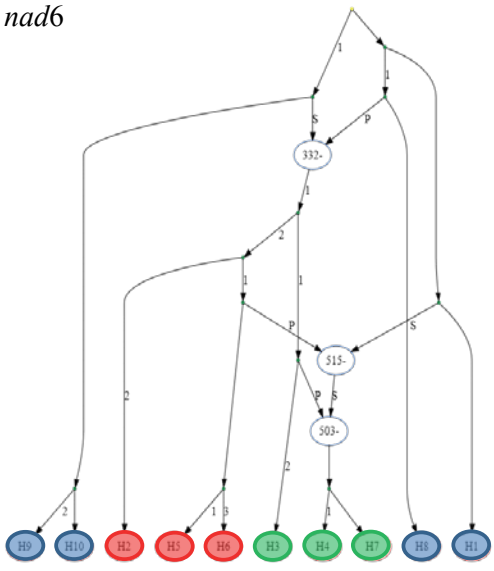
nad5



H1 = *F. circinatum*
H2 = *F. nygamai*
H3 = *F. verticillioides*
H4 = *F. thapsinum*
H5 = *F. thapsinum*
H6 = *F. mangiferae*
H7 = *F. temperatum*
H8 = *F. mangiferae*
H9 = *F. strelihyphosum*
H10 = *F. fujikuroi*
H11 = *F. proliferatum*
H12 = *F. fujikuroi*
H13 = *F. circinatum*
H14 = *F. sacchari*
H15 = *F. subglutinans, F. temperatum*
F. konzum
H16 = *F. strelihyphosum*

= American Clade
= African Clade
= Asian Clade

nad6



H1 = *F. circinatum*
H2 = *F. verticillioides*
H3 = *F. sacchari*
H4 = *F. fujikuroi*
H5 = *F. nygamai*
H6 = *F. thapsinum*
H7 = *F. mangiferae, F. proliferatum, F. fujikuroi*
H8 = *F. konzum*
H9 = *F. subglutinans, F. temperatum*
H10 = *F. strelihyphosum*

= American Clade
= African Clade
= Asian Clade

from ancestral individuals that became haplotype H5 and H6 and therefore haplotypes that represent both the "African" and "Asian" clades (Fig. 2).

Evaluating possible sources of systematic error and/or false positives

To evaluate the effect of systematic error on our coalescent-based analyses, various additional parameters were estimated and examined. This is because failure to appropriately account for the complex properties of the individual gene datasets could lead to false detection or non-detection of ancestral recombination events. In other words, these analyses provided an indication of the robustness of the conclusions drawn from the RECMIN and ARG results. Indeed, based on our analyses, neither nucleotide diversity nor substitution saturation (homoplasy) appeared to represent significant sources of such systematic errors. The average nucleotide diversity (π) estimated for all of the mt datasets (see Table 1 for EMBL nucleotide sequence database numbers) was low and ranged from 0.4–1 % (Table 4). For these datasets, the sequence divergence estimates D_{xy} and D_a between FFSC and *F. oxysporum* were also low and ranged between 0.9–2 %, and 0.3–1.5 %, respectively (Table 4). In terms of substitution saturation, the observed entropy I_{SS} was compared to $I_{SS.C}$. For the five gene datasets used in this study, as well as for the datasets respectively containing first plus second codon position and third codon positions only, the I_{SS} values were significantly smaller than the $I_{SS.C}$ values (Table 5), which suggested negligible substitution saturation.

Selection analyses with FEL, REL and SLAC suggested that no codons were under positive selection. In contrast, the FEL, REL and SLAC analyses identified a number of codons under negative/purifying selection for all the datasets included (Supplementary Table S4). For example, 13, 10, 6, 31 and 20 codons of the *atp*6, *cox*2, *nad*3, *nad*5, and *nad*6 gene datasets, respectively, were identified to be under negative selection by one and/or two of the methods used. The consensus results, however, suggested that only *nad*5 and *nad*6, respectively, had three [(codon 139; methionine), (287; phenylalanine), (356; leucine)] and one (codon 116; serine) codon under negative selection (significant p-values FEL and SLAC = 0.1; REL baysian factor = 50). However, none of the consensus codons in the *nad*5 and *nad*6 genes were shown to be subjected to negative selection overlapped with the recombination events suggested from the ARGs.

The only potential source of systematic error in our coalescent-based analyses was among site rate heterogeneity. The gamma distribution shape parameter varied among the different mt gene datasets (Table 4), but was particularly low for the *atp*6 and *cox*2 datasets (i.e. α-values of 0.01 as opposed to ³ 0.3 for the other three datasets). This indicated strong substitution rate variation among sites that could lead to the detection of false positive recombination events. The RECMIN and ARG results did not predict any recombination events in the *cox2* dataset, but the single recombination event predicted in the *atp6* dataset likely represents an analytical artefact (Posada 2002).

DISCUSSION

This aim of this study was to find evidence of heteroplasmy-associated recombination between mt genomes of species in the FFSC. Both direct and coalescent-based method allowed for the detection of heteroplasmy-associated recombination, which support previous suggestions that gene flow or introgressive hybridization occurred in the history of the FFSC (O'Donnel & Cigelnik 1997, Fourie *et al.* 2013). In addition, the detection of recombination by both methods also provided evidence that introgressive hybridization occurred at ancient and more recent time scales.

Both of the approaches used in this study allowed for the identification of interspecies recombination events in the ancestry of the FFSC. The methods implemented in the RDP3 package infer recombination events directly from the sequence information provided (Martin *et al*. 2010), while the ARGs provide information on the order in which recombination and mutation occurred over evolutionary time (Lyngsø *et al*. 2005). In other words, ARGs represent statistical descriptions of the genealogical history of each mt gene sequences backwards in time to the most recent common ancestor (Griffiths 1999, McVean *et al*. 2002, Lyngsø *et al*. 2005). Although examples in fungi are limited, RDP3's direct recombination detection methods detected recombination and hybridization between free-ranging Australian lizards (Ujvari *et al*. 2007), between

Table 4. Gene diversity, divergence and rate heterogeneity of the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) examined in this study.

Gene	Diversity and Divergence[a]						Rate heterogeneity[b]
	π	π (%)	D_{xy}	D_{xy} (%)	D_a	D_a (%)	
*atp*6	0.00569	0.5	0.02032	2	0.01428	1	0.01
*nad*3	0.00934	0.9	0.01116	1	0.00318	0.3	0.58
*nad*5	0.00429	0.4	0.0102	1	0.00696	0.6	0.30
*nad*6	0.01015	1	0.00967	0.9	0.0046	0.4	0.46
*cox*2	0.00627	0.6	0.01882	1.8	0.01569	1.5	0.01

[a]Sequence diversity (Nei 1987) and divergence (Nei 1987) for each dataset were determined using DNAsp ver. 5 (Librado & Rozas 2009). π = the average number of nucleotide differences per site between two sequences, D_{xy} = the average number of nucleotide substitutions per site between populations, D_a = the number of net nucleotide substitutions per site between populations. π, D_{xy}, and D_a values were converted to percentages for comparative purposes.

[b]The shape of the gamma distribution which indicate the pattern of among-site rate heterogeneity for all gene datasets were inferred by with jModeltest (Posada 2008).

Table 5. Index of substitution saturation as well as the critical value of the index of substitution saturation used to measure the level of substitution saturation within the FFSC mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6, and *cox*2).

Genes	Saturation (1+2 position)		Saturation (3 position)	
	Iss.critical[b]	Iss[a]	Iss.critical[b]	Iss[a]
*atp*6	0.7009	0.0173	0.6111	0.0197
*nad*3	0.7341	0.0102	0.7922	0.0341
*nad*5	0.7758	0.0017	0.7283	0.0263
*nad*6	0.6941	0.0043	0.6429	0.045
*cox*2	0.6886	0.1582	0.6886	0.2665

[a]Iss = information entropy-based index of substitution saturation determined with DAMBE5 (Xia 2013) by testing if the observed entropy at site *i* is significantly smaller than the expected entropy under full substitution saturation.

[b]Iss.critical = critical I_{SS} value estimated with DAMBE 5 (Xia 2013). The critical I_{SS} depend on the topology of the true tree, the number of OTUs, the sequence length, nucleotide frequency and the transition/transversion ratio (Xia *et al*. 2003, Xia & Lemey 2009).

scorpion species in the family *Buthidae* (Gantenbein *et al*. 2005) and between divergent populations of the nematode species *Globodera pallida* (Hoolahan *et al*. 2012), while ARG analysis showed introgression or hybridization in organisms such hydrothermal vent mussels (Faure *et al.* 2009).

We performed extensive analyses to ensure that the putative ancestral recombination events detected in this study were not due to a failure to account for the inherent evolutionary complexity of the data (Possada 2002, Delsuc *et al.* 2005). Despite nucleotide substitution saturation possibly being common in mt genomes (Spencer & Barral 2012, Gaillardin *et al*. 2012), little evidence for it was found as was expected since the FFSC mt datasets were highly conserved. However, as previously described (Dowling *et al*. 2008, Rand 2001, Stewart *et al*. 2008, Soares *et al*. 2009), the various FFSC mt gene datasets contained evidence of selection at specific codons. But we only detected purifying selection (codons under positive or diversifying selection were not detected) and none of the affected consensus sites (i.e. identified by all methods) occurred in the recombinant regions identified. Also as expected (Excoffier & Yang 1999, Ingman *et al*. 2000), strong among-site rate variation was detected in some of the mt genes examined, which may be linked to the function/structure of their products (Yang 1996). Although the infinite-sites model of substitution (Kimura 1969) utilized in the ARG analyses likely excluded the effects of this phenomenon, only one putative recombination event was detected in a dataset associated with strong among-site rate variation. Taken together, these results thus indicated that the putative recombination events identified do not represent analytical artefacts, because the sources of systematic error in the various FFSC mt gene datasets were limited.

One concern that could not be fully eliminated in the current study is that the DNA sequence signatures of introgression or hybridization are not readily distinguishable from those of incomplete lineage sorting or deep coalescence (Maddison 1997, Degnan & Rosenberg 2009). Incomplete lineage sorting typically manifests as polymorphisms that persist through several speciation events; i.e. divergence and drift-associated random sorting of an ancestral polymorphism did not lead to its differential fixation in the resulting species (Maddison 1997). In ancestral recombination analyses, incompletely sorted polymorphisms would thus "behave" in a similar manner to those originating from interspecies gene flow (Degnan & Rosenberg 2009). Although we could not rule out its involvement in our analyses, incomplete lineage sorting is unlikely to have affected all of the recombinant sites/regions identified. This is especially true for the long stretches of recombinant sequences (199–1668 bp) detected among the genomes of the "Asian" and "American" clade species included (see Table 3). Future studies should however investigate the role of incomplete lineage sorting in the evolution of FFSC by employing statistical approaches to distinguish gene flow and incomplete lineage sorting based on whole genome sequence data (Joly *et al*. 2009).

Mechanisms that would allow for recombination between mt genomes of different FFSC species is unknown. It is currently hypothesized that recombination between different mt genomes can occur via the dispersed repeat elements they harbour, exchange between highly conserved regions or *via* intron homing (Basse 2010, Galtier 2011). Recombination between dispersed repeat elements is common among plant mitochondria in which the repeats serve as crossover points for homologous recombination (Palmer & Herbon 1988). Recombination via intron homing occurs when LAGLIDADG or GIY-YIG endonucleases that are encoded in fungal mitochondrial introns move into previously intron-less genes (Goddard & Burt 1999, Haugen *et al*. 2005, Stoddard 2006). Given the overall gene order conservation of the intron-rich mt genomes of the FFSC described here and previously (Al-Reedy *et al*. 2012, Fourie *et al*. 2013), recombination *via* intron homing and/or exchange between conserved regions is potentially more likely as recombination *via* dispersed repeats would allow for gene order rearrangements. In general, however, mt recombination in fungi is expected to employ mechanisms that are markedly different from those inferred for animals where mt genomes typically lack introns and dispersed repetitive elements (Rokas *et al.* 2003, Piganeau *et al*. 2004).

Overall, the findings presented here indicate that interspecies or heteroplasmy-associated gene flow and recombination occurred at both ancient and recent timescales during the evolution of the FFSC. The results of the ARG analyses presented here (especially the *nad*5 and *nad*6 ARGs) provide evidence for older and/or ancient recombination within the FFSC. It is conceivable that such recombination and subsequent introgression events could have occurred in the Miocene, prior to biogeographic separation of the clades

(O'Donnell *et al.* 2013), during which diversification of the complex coincided with the radiation of grasses and eudicots that use C_4 photosynthesis to fix carbon (Edwards *et al*. 2010, Christin *et al.* 2011, O'Donnell *et al.* 2013). During this time period, environmental conditions likely influenced the distribution of the ancestral FFSC members, thus providing the opportunity for introgressive hybridization to occur (Olson & Stenlid 2002, Schardl & Craven 2003).

Recombination events that occurred at more recent timescales were revealed by the tools implemented in RDP3 (Martin *et al.* 2010). These all detect recombination by identifying regions of sequence similarity between individuals that are unusually high in comparison to the overall sequence similarity of these individuals as estimated from the entire region and/or genome in question. However, post-recombination mutations (i.e. those that accumulate over evolutionary time after the original interspecies gene flow event) would obscure the distinction of these recombinant regions from other background mutations (Posada 2002). Accordingly, recombinant regions detected by these methods most likely represent sites at which the signatures of recombination have not yet been eroded away by normal mutational processes. The five significant regions of recombination detected here (see Table 3), were thus the result of interspecies gene flow events that occurred relatively recently in the history of the FFSC, although information regarding the geographic contact points of these events are lacking. Also, many FSSC species are inter-fertile under laboratory conditions (Desjardins *et al.* 2000, Leslie *et al.* 2004a) and a natural hybrid has been described from native tall-grass prairie south of Manhattan in Kansas (Leslie *et al.* 2004b, 2007). Overall the results of this study showed that interspecies gene flow and introgressive hybridization have played an important role in the evolution of the FFSC and will likely continue to do so. However, the extent to which these phenomena would influence the evolution of the complex and at what point new species will emerge remains to be determined.

ACKNOWLEDGMENTS

This work is based on research supported in part by a number of grants from the South African National Department of Science and Technology (DST), National Research Foundation (NRF), and the Technology and Human Resources of Industry Programme (THRIP) (includes Grant specific unique reference number (UID) 83924). The grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by NRF supported research are that of the author(s) and that the NRF accepts no liability whatsoever in this regard. We further acknowledge the following programmes and institutions for financial support: the Tree Protection Cooperative Programme (TPCP),L'Oréal/UNESCO for Women in Science in Sub-Saharan Africa, The Claude Leon foundation, and the University of Pretoria.

REFERENCES

Al-Reedy RM, Malireddym R, Dillman CB, Kennell JC (2012) Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genetics and Biology* **49**: 2–14.

Aylor DL, Price EW, Carbone I (2006) SNAP, combine and map modules for multilocus population genetic analysis. *Bioinformatics* **22**: 1399–1401.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**: 729–744.

Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**: 39–50.

Basse CW (2010) Mitochondrial inheritance in fungi. *Current Opinion in Microbiology* **13**: 1–8.

Brasier CM, Kirk SA, Pipe ND, Buck KW (1998) Rare interspecific hybrids in natural populations of the Dutch elm disease pathogen *Ophiostoma ulmi* and *O. novo-ulmi*. *Mycological Research* **102**: 45–57.

Christin P-A, Osborne CP, Sage RF, Arakaki M, Edwards EJ (2011) C4: eudicots are not younger than C4 monocots. *Jounal of Experimental Botany* **62**: 3171–3181.

Cunnington JH (2007) Organization of the mitochondrial genome of *Fusarium oxysporum* (anamorphic *Hypocreales*). *Mycoscience* **48**: 403–406.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24**: 332–340.

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* **6**: 361–375.

Desjardins AE, Plattner RD, Gordon TR (2000) *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. *Mycological Research* **104**: 856–872.

Dowling DK, Friberg U, Lindell J (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology and Evolution* **23**: 546–554.

Edwards EJ, Osborne CP, Strömberg AE, Smith S (2010) The origins of C4 grasslands, integrating evolutionary and ecosystem science. *Science* **328**: 578–591.

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Molecular Biology and Evolution* **16**: 1357–1368.

Eyre-Walker A, Smith NH, Smith J M (1999) How clonal are human mitochondria? *Proceedings of the Royal Society of London, B Biological Science* **266**: 477–483.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent Mussels. *PLOS One* **4**: e6485.

Fourie G, Van der Merwe NA, Wingfield BD, Bogale M, Tudzynski B, Wingfield MJ, Steenkamp ET (2013) Evidence for inter-specific recombination among the mitochondrial genomes of Fusarium species in the *Gibberella fujikuroi* complex. *BMC Genomics* **14**: 605.

Gaillardin C, Neuveglise C, Kerscher S, Nicaud J-M (2012) Mitochondrial genomes of yeasts of the Yarrowia clade. *FEMS Yeast Research* **12**: 317–331.

Galtier N (2011) The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology* **9**: 61–63.

Galtier N, Enard D, Radondy Y, Bazin E, Belkhital K (2006) Mutation hot spots in mammalian mitochondrial DNA. *Genome Research* **16**: 215–222.

Gantenbein B, Fet V, Gantenbein-Ritter IA, Balloux F (2005) Evidence for recombinationin scorpion mitochondrial DNA. *Proceedings of the Royal Society of London* **272**: 679–704.

Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharya MK, *et al*. (2013) One Fungus, one name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400–408.

Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Acadamy of Sciences, USA* **96**: 13880–13885.

Griffiths RC (1999) The time to the ancestor along sequences with recombination. *Theoretical Population Biology* **55**: 137–144.

Groenewald S, Van Den Berg N, Marasas WFO, Viljoen A (2006) The application of high-throughput AFLPs in assessing genetic diversity in *Fusarium oxysporum* f.sp. *cubense*. *Mycological Research* **110**: 297–305.

Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. *Trends in Genetics* **19**: 84–90.

Hartley JL, Bowen H (2003) PEG precipitation for selective removal of small DNA fragments. *Focus* **25**: 18.

Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends in Genetics* **21**: 111–119.

Huang D, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the Metazoan tree and Its implications for DNA barcoding. *Journal of Molecular Evolution* **66**: 167–174.

Hoolahan AH, Blok VC, Gibson G, Dowton D (2012) Evidence of animal mtDNA recombination between divergent populations of the potato cyst nematode *Globodera pallida*. *Genetica* **140**: 19–29.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–165.

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708–713.

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics* **171**: 1951–1962.

Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* **174**: E54–E70.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics and Breeding* **61**: 893–903.

Kosakovsky Pond SL, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.

Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.

Kvas M, Marasas WFO, Wingfield BD, Wingfield MJ, Steenkamp ET (2009) Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. *Fungal Diversity* **34**: 1–21.

Lang FB, Laforest M-J, Burger G (2007) Mitochondrial introns: a critical view. *Trends in Genetics* **23**: 119–125.

Leslie JF, Anderson LL, Bowden RL, Lee Y-W (2007) Inter-and intra-specific genetic variation in *Fusarium. International Journalof Food Microbiology* **119**: 25–32.

Leslie JF, Summerell BA (2006) *The Fusarium Laboratory Manual*. Oxford: Blackwell Publishing.

Leslie JF, Zeller KA, Logrieco A, Mule G, Moretti A, Ritieni A (2004a) Species diversity of and toxin production by *Gibberella fujikuroi* species complex strains isolated from native prairie grasses in Kansas. *Applied and Environmental Microbiology* **70**: 2254–2262.

Leslie JF, Zeller KA, Wohler M, Summerell BA (2004b) Interfertility of two mating populations in the *Gibberella fujikuroi* species complex. *European Journal of Plant Patholology* **110**: 611-619.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955–964.

Lyngsø RB, Song YS, Hein J (2005) Minimum recombination histories by branch and bound. *Proceedings of the 5th International Workshop on Algorithms in Bioinformatics, Lecture Notes in Bioinformatics* **3692**: 239–250.

Ma L-J, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, *et al*. (2010) Comparative genomics reveals mobile pathogenicity chromosome in *Fusarium. Nature* **464**: 356–373.

Maddison WP (1997) Gene trees in species trees. *Systems Biology* **46**: 523–536.

Mallet J (2007) Hybrid speciation. *Nature Reviews* **446**: 279–283.

Marechal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299–317.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**: 562-563.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**: 2462–2463.

Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified BOOTSCAN algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses* **21**: 98–102.

Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**: 126–129.

McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

Myers SR, Griffiths RC (2002) Bounds on the minimum number of recombination events in a sample history. *Genetics* **163**: 375–394.

Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Reviewof Genetics* **39**: 197–218.

O'Donnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Molecular Phylogenetics and Evolution* **7**: 103–116.

O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematic and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* **90**: 465–493.

O'Donnell K, Kistlet HC, Tacke BK, Casper HH (2000a) Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proceedings of the National Acadamy of Sciences, USA* **97**: 7905–7910.

O'Donnell K, Nirenberg HI, Aoki T, Cigelnik E (2000b) A multigene phylogeny of the *Gibberella fujikuroi* species complex, Detection of additional phylogenetically distinct species. *Mycoscience* **41**: 61–78.

O'Donnell K, Rooney AP, Proctor RH, Brown DW, McCormick SP, *et al.* (2013) Phylogenetic analyses of *RPB*1 and *RPB*2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genetics and Biology* **52**: 20–31.

Olson A, Stenlid J (2002) Pathogenic fungal species hybrids infecting plants. *Microbes and Infection* **4**: 1353–1359.

Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.

Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Current Opinion in Genetics and Development* **14**: 675–679.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**: 218–225.

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* **28**: 87–97.

Pantou MP, Kouvelis VN, Typas MA (2008) The complete mitochondrial genome of *Fusarium oxysporum*, insights into fungal mitochondrial evolution. *Gene* **419**: 7–15.

Piganeau GI, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution* **21**: 2319–2325.

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences, empirical data. *Molecular Biology and Evolution* **19**: 708–717.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**: 1253–1256.

Price EW, Carbone I (2005) SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics* **21**: 402–404.

Rand DM (2001) The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* **32**: 415–448.

Reed FA, Tishkoff SA (2006) Positive selection can create false hotspots of recombination. *Genetics* **172**: 2011–2014.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution* **18**: 411–417.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology* **61**: 225–247.

Rubinoff D, Holland BS (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54**: 952–961.

Schardl CL, Craven KD (2003) Interspecific hybridization in plant-associated fungi and oomycetes, a review. *Molecular Ecology* **12**: 2861–2873.

Seifert KA, Samson RA, DeWaard JD, Houbraken J, Levesque CA, *et al.* (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Acadamy of Sciences, USA* **104**: 3901–3906**.**

Soares P, Ermini L, Thomson N, Mormina M, Rito T, *et al.* (2009) Correcting for purifying selection, an improved human mitochondrial molecular clock. *American Journal of Human Genetics* **84**: 740–759.

Spencer PS, Barral JM (2012) Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology* **1**: 1–8.

Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, *et al.* (2008) Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biology* **6**: 63–71.

Stoddard BL (2006) Homing endonuclease structure and function. *Quarterly Reviews of Biophysics* **38**: 49–95.

Taylor JW (1986) Fungal evolutionary biology and mitochondrial DNA. *Experimental Mycology* **10**: 259–269.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution* **22**: 925–933.

Turner E, Jacobson DJ, Taylor JW (2010) Reinforced postmating reproductive isolation barriers in *Neurospora*, an ascomycete microfungus. *Journal of Evolutionary Biology* **23**: 1642–1656.

Ujvari B, Dowton M, Madsen T (2007) Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biology Letters* **3**: 189–192.

Wilson CC, Bernatchez L (1998) The ghost of hybrids past, fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology* **7**: 127–132.

Wiuf C, Christensen T, Hein J (2001) Simulations study of the reliability of recombination detection methods. *Molecular Biology and Evolution* **18**: 1929–1939.

Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **30**: 1720–1728.

Xia X, Lemey P (2009) Assesing substitution saturation with DAMBE. In: *The Phylogenetic Handbook, a practical approach to phylogenetic analysis and hypothesis testing* (Salemi M, Lemey P, Vandamme AM eds): 611–626. Cambridge: Cambridge University Press.

Xia X, Xie Z, Salemi M, Chen M, Wang Y (2003) An index of substitution saturarion and its application. *Molecular Phylogenetics and Evolution* **26**: 1–7.

Yang Z (1996) Among-site rate variation and its impact on phylogenetica nalyses. *Trends in Ecology and Evolution* **11**: 367–372.

Mitochondrial introgression and interspecies recombination in the *Fusarium fujikuroi* species complex.

Gerda Fourie[1], Nicolaas A. van der Merwe[2], Brenda D.Wingfield[2], Mesfin Bogale[2], Michael J Wingfield[1] and Emma T Steenkamp[1].

Departments of Microbiology and Plant Pathology[1] and Genetics[2], Forestry and Agricultural Biotechnology Institute (FABI)[3], University of Pretoria, Pretoria, South Africa.
*gerda1.fourie@fabi.up.ac.za.

Table S1. Primer sequences and annealing temperatures of the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) designed in this study to amplify and sequence species within the *Fusarium fujikuroi* species complex.

Table S2. The recombinant regions detected by RDP, Geneconv, Bootscan and Maximum χ^2 of the *Fusarium fujikuroi* mitochondrial genomes.

Table S3. Individual haplotypes for the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) used to construct the minimum number of recombination events as well as minimal ancestral recombination graphs in SNAP workbench.

Table S4. Haplotype matrix and the BEAGLE branch and bound log for the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) used to construct the minimum number of recombination events as well as minimal ancestral recombination graphs in SNAP workbench.

Table S5. Sites under negative selection identified with the Single Likelihood Ancestor Counting (SLAC) method, the Fixed Effect Likelihood (FEL) method, and the Random Effect Likelihood (REL) method from Datamonkey (http://www.datamonkey.org/) among the mitochondrial genes used in this study.

Table S1. Primer sequences and annealing temperatures of the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) designed in this study to amplify and sequence species within the *Fusarium fujikuroi* species complex.

Gene	Annealing temperature	Sequence 5′ → 3′
*nad*3	54°C	AGT GTA ACT TTT CTT TTT GTT TTT GTT
		TTC AAG GCA TTT TTA CCT AAT TCA
*nad*5	55°C	TTT TTG GTA GAA AAG TCG GTG TT
		CGT CAT CTA CTA CCT GAA CAA AAG
nad6	57°C	CCG AGA TTT TCT CGA ATG GA
		AGT AAT TAT TAT TGC ACC TAC CAT AGC
*cox*2	55°C	TGA TGC ACC TAT GCC TTG AG
		TTT CTG AAC ATT GCC CAT AGA A
*atp*6	55°C	GCG CAA GCT GAA GAA ATA AGA
		CCT GGA TGA AGG CTA TTG CT

Table S2. The recombinant regions detected by RDP, Geneconv, Bootscan and Maximum χ^2 of the *Fusarium fujikuroi* mitochondrial genomes.

Recombinant[a]	Recombinant region[b]			Parents[c]	P-values [d]			
	Begin	End	Genes included		RDP	GeneConv	Bootscan	Maxχ^2
F. circinatum	20 366	20934	intergenic between tRNA.cys and *cox*1	*F.mang* X unknown	6.89E-04	3.96E-02	4.54E-03	6.25E-05
F. circinatum	38743	42640	*nad*4 to rns	*F.fuj* X unknown	2.30E-03	NS	NS	1.60E-03
F. fujikuroi	3075	3196	*atp*9	*F.mang* X unknown	NS	1.37E-03	NS	1.22E-02
F. fujikuroi	9819	10355	*cob*	*F.cir* X unknown	NS	3.09E-02	NS	NS
F. fujikuroi	13525	13689	intergenic between tRNA.cys and cox1	*F.mang* X unknown	NS	1.10E-04	3.26E-06	7.97E-03
F. fujikuroi	15721	15945	intergenic between *cox*1 and tRNA.arg	*F.cir* X unknown	NS	4.04E-09	1.72E-03	1.16E-03
F. fujikuroi	23688	23777	ribosomal small subunit	*F.mang* X unknown	NS	3.03E-02	NS	NS
F. mangiferae	1047	1673	*nad*2	*F.fuj* X *F.ver*	NS	1.13E-02	NS	NS
F. mangiferae	2862	4530	*atp*9 to *cox*2	*F.fuj* X *F.cir*	2.50E-45	2.29E-48	2.46E-46	8.77E-19
F. mangiferae	7974	8265	intergenic between *nad*5 and tRNA.arg	*F.cir* X unknown	NS	NS	NS	3.39E-03
F. mangiferae	10621	10820	intergenicbetween*cob*and tRNA.cys	*F.fuj* X unknown	1.03E-13	2.59E-11	8.52E-13	4.14E-05
F. mangiferae	14267	17524	*cox*1	*F.fuj* X unknown	9.83E-07	NS	NS	2.27E-03
F. mangiferae	17675	18259	intergenic between cox1 and tRNA.arg	*F.fuj* X *F.cir*	NS	3.34E-40	NS	3.62E-13
F. mangiferae	23296	24471	*atp*6	*F.fuj* X *F.cir*	3.90E-11	5.08E-08	2.29E-11	1.16E-06
F. mangiferae	117	749	*nad*2	*F.cir* X unknown	NS	1.30E-05	NS	NS
F. mangiferae	18436	21436	*nad*1	*F.fuj* X unknown	1.43E-09	NS	NS	3.66E-03
F. mangiferae	4620	4767	intergenicbetwee*cox*2 and tRNA.arg	*F.fuj* X unknown	NS	4.34E-03	NS	NS
F. temperatum	11143	11771	tRNA.arg	*F.mang* X unknown	1.78E-22	2.54E-24	9.66E-27	2.47E-11
F. temperatum	13002	15256	*cob*	*F.cir* X *F.ver*	NS	5.79E-03	NS	NS
F. temperatum	33095	33353	*nad*1	*F.cir* X *F.mang*	NS	8.72E-03	NS	NS
F. temperatum	34713	34925	*nad*4	*F.cir* X unknown	2.52E-27	9.19E-29	NS	9.49E-08
F. verticillioides	27354	27494	intergenic between *atp6* and rns	*F.cir* X unknown	5.11E-04	1.89E-05	NS	4.40E-03

[a] The FFSC species or daughter in which the recombinant region was detected.

[b] The position of the recombinant region according to the mt genome of each specific FFSC species, *F. mangiferae* (KP742837), *F. temperatum* (KP742838) and *F. circinatum* (JX910419).

[c] The suggested origin of the recombinant region according to RDP, Geneconv, Bootscan and Maximum χ^2 implemented in RDP version 3.44 (Martin et al. 2010). Unknown = the parent of this recombination event is not present within the dataset.

[d] The P-values for each of the recombination detection methods RDP, Geneconv, Bootscan and Maximum χ^2 implemented in RDP version 3.44 (Martin et al. 2010). RDP identifies recombinant segments via pair-wise percentage identity values (Martin & Rybicki, 2000), Geneconv detects recombinant segments as aligned pairs that are unusually long and sufficiently similar (Padidam et al. 1999), Bootscan identifies recombinant segments as high degree of bootstrap support for different topologies (Martin et al. 2005) and Maximum χ^2 detects recombinant sections by comparing the number of segregating sites on both sides of the recombinant breakpoint (Maynard Smith, 1992). NS = non-significant p-value.

Table S3. Individual haplotypes for the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) used to construct the minimum number of recombination events as well as minimal ancestral recombination graphs in SNAP workbench.

*atp*6	
H1	CMWF350, CMWF498
H2	CMWF497
H3	CMWF1209, CMWF1210
H4	CMWF1215, CMWF1216
H5	NRRL29056, CMWF1196, CMWF1197
H6	CMWF1220, CMWF1200, CMWF1201
H7	CMWF1213, CMWF1214, CMWF1198, CMWF1199, CMWF1202, CMWF1203
H8	CMWF1207, CMWF1208
H9	CMWF1204, CMWF1205, CMWF1206, CMWF398
H10	CMWF1211, CMWF1212

*nad*3	
H1	CMWF350, CMWF497, CMWF498, CMWF1204, CMWF1205, CMWF1206, CMWF389, CMWF1211, CMWF1212, CMWF1215, CMWF1216
H2	NRL29056, CMWF1196, CMWF1197
H3	CMWF1220, CMWF1200, CMWF1201, CMWF1213, CMWF1214, CMWF1202, CMWF1203
H4	CMWF1198, CMWF1199
H5	CMWF1207, CMWF1208
H6	CMWF1209, CMWF1210

*nad*5	
H1	CMWF350, CMWF497
H2	CMWF1209, CMWF1210
H3	NRRL29056, CMWF1196, CMWF1197
H4	CMWF1207
H5	CMWF1208
H6	CMWF1214
H7	CMWF389
H8	CMWF1213

H9	CMWF1215
H10	CMWF1220
H11	CMWF1202, CMWF1203
H12	CMWF1200, CMWF1201
H13	CMWF498
H14	CMWF1198, CMWF1199
H15	CMWF1204, CMWF1205, CMWF1206, CMWF1211, CMWF1212
H16	CMWF1216

*nad*6

H1	CMWF350, CMWF479, CMWF498
H2	NRL29056, CMWF1196, CMWF1197
H3	CMWF1198, CMWF1199
H4	CMWF1220
H5	CMWF1209, CMWF1210
H6	CMWF1207, CMWF1208
H7	CMWF1200, CMWF1201, CMWF1213, CMWF1214, CMWF1202, CMWF1203
H8	CMWF1211, CMWF1212
H9	CMWF1204, CMWF1205, CMWF1206, CMWF389
H10	CMWF1215, CMWF1216

*cox*2

H1	CMWF350, CMWF497, CMWF498, CMWF1204, CMWF1205, CMWF1206, CMWF389, CMWF1211, CMWF1212, CMWF1215, CMWF1216
H2	CMWF1207, CMWF1208
H3	CMWF1198, CMWF1199
H4	CMWF1196, CMWF1197, NRRL29056
H5	CMWF1220, CMWF1200, CMWF1201, CMWF1213, CMWF1214, CMWF1202, CMWF1203
H6	CMWF1209, CMWF1210

Table S4. Haplotype matrix and the BEAGLE branch and bound log for the mitochondrial genes (*nad*3, *nad*5, *nad*6, *atp*6 and *cox*2) used to construct the minimum number of recombination events as well as minimal ancestral recombination graphs in SNAP workbench.

*atp*6

```
Position           12234455
                  5643640827
                  2580947258

Site Number              1
                  1234567890

Consensus         0010100110
Character Type    i-i-ii-ii-
H1          (  2) ..0.0..00.
H2          (  1) 1.0.0..00.
H3          (  2) 1....1....
H4          (  2) .10.0..0..
H5          (  3) .....1....
H6          (  3) ......1...
H7          (  6) ..........
H8          (  2) .....1...1
H9          (  4) ..010..0..
H10         (  2) ..0.0..0..
```

Mutation of site 2 in sequence H4
Mutation of site 4 in sequence H9
Mutation of site 7 in sequence H6
Mutation of site 10 in sequence H8
Coalescing sequences H6 and H7
Coalescing sequences H5 and H8
Coalescing sequences H4 and H9
Coalescing sequences H4 and H10
Recombination in sequence H2 after site 2; prefix is new sequence 11
Coalescing sequences H3 and 11
Mutation of site 1 in sequence H3
Coalescing sequences H1 and H2
Coalescing sequences H3 and H5
Mutation of site 6 in sequence H3
Coalescing sequences H3 and H6
Mutation of site 3 in sequence H3
Mutation of site 5 in sequence H3
Mutation of site 8 in sequence H3
Coalescing sequences H3 and H4
Mutation of site 9 in sequence H3
Coalescing sequences H1 and H3

*nad*3

```
Position        11222
                55619338
```

```
            24978455

Site Number 12345678

Consensus   10000100
Character Type  -ii---i-
H1      ( 11) 0....0..
H2      (  3) .......1
H3      (  7) .11...1.
H4      (  2) .11.1.1.
H5      (  2) ...1....
H6      (  2) ........
```

Mutation of site 4 in sequence H5

Mutation of site 5 in sequence H4

Mutation of site 8 in sequence H2

Coalescing sequences H3 and H4

Coalescing sequences H2 and H5

Coalescing sequences H2 and H6

Mutation of site 2 in sequence H3

Mutation of site 3 in sequence H3

Mutation of site 7 in sequence H3

Coalescing sequences H2 and H3

Mutation of site 1 in sequence H2

Mutation of site 6 in sequence H2

Coalescing sequences H1 and H2

*nad*5

```
Position                        11111111111111111111
                    12223344456899901111222344455555667
                    30661467784203920135267965560088812 1
                    10490251778590967616418655841678919 9

Site Number                  11111111112222222223333333
                    12345678901234567890123456789012345 6

Consensus           00000000000010010010001100100000000 1
Character Type      -i----i-i-iii-iiiii-i-ii--ii-i----ii
H1          (  2) ............0..0..0...00..0........0
H2          (  2) 11............1.1.......1.........0
H3          (  3) .1.1.1........1011..11...1.......1.0
H4          (  1) .1............1.1...1............1.
H5          (  1) .1............1.1...1............10
H6          (  1) ..1...1.1...0......1.......1........
H7          (  1) ....1..1.11....0..0...00..0.........
H8          (  1) ......1.1..................1........
H9          (  1) ......1.1..1..............1........
H10         (  1) ........1..1..............1.1111...
H11         (  2) ........1..1.....1.........1........
H12         (  2) ........1..1..............1.1......
H13         (  1) ..........1.0..0..0...00..0........0
```

```
H14         (  2) ..........1.1.............1.......
H15         (  5) ..............0..0...00..0........0
H16         (  1) ..............0..0...00...........0
```

Mutation of site 1 in sequence H2
Mutation of site 3 in sequence H6
Mutation of site 4 in sequence H3
Mutation of site 5 in sequence H7
Mutation of site 6 in sequence H3
Mutation of site 8 in sequence H7
Mutation of site 10 in sequence H7
Mutation of site 14 in sequence H14
Mutation of site 20 in sequence H6
Mutation of site 22 in sequence H3
Mutation of site 25 in sequence H2
Mutation of site 26 in sequence H3
Mutation of site 29 in sequence H14
Mutation of site 31 in sequence H10
Mutation of site 32 in sequence H10
Mutation of site 33 in sequence H10
Mutation of site 34 in sequence H3
Coalescing sequences H10 and H12
Mutation of site 30 in sequence H10
Recombination in sequence H3 after site 17; suffix is new sequence 17
Recombination in sequence 17 after site 20; suffix is new sequence 18
Coalescing sequences H11 and 17
Mutation of site 18 in sequence H11
Coalescing sequences H10 and H11
Recombination in sequence H4 after site 35; prefix is new sequence 19
Coalescing sequences H5 and 19
Mutation of site 35 in sequence H5
Coalescing sequences H6 and H4
Coalescing sequences H5 and 18
Mutation of site 21 in sequence H5
Coalescing sequences H2 and H5
Recombination in sequence H2 after site 15; prefix is new sequence 20
Coalescing sequences H3 and 20
Mutation of site 2 in sequence H3
Mutation of site 15 in sequence H3
Recombination in sequence H3 after site 16; prefix is new sequence 21
Coalescing sequences H15 and 21
Coalescing sequences H2 and H3
Mutation of site 17 in sequence H2
Recombination in sequence H7 after site 12; prefix is new sequence 22
Coalescing sequences H13 and 22
Mutation of site 11 in sequence H13
Coalescing sequences H1 and H13
Recombination in sequence H8 after site 12; prefix is new sequence 23
Coalescing sequences H6 and 23
Coalescing sequences H9 and H8

Recombination in sequence H6 after site 11; prefix is new sequence 24
Coalescing sequences H9 and 24
Mutation of site 7 in sequence H9
Coalescing sequences H9 and H10
Mutation of site 9 in sequence H9
Recombination in sequence H6 after site 15; prefix is new sequence 25
Coalescing sequences H1 and 25
Coalescing sequences H9 and H6
Mutation of site 28 in sequence H9
Coalescing sequences H9 and H14
Mutation of site 12 in sequence H9
Recombination in sequence H7 after site 35; prefix is new sequence 26
Coalescing sequences H15 and 26
Coalescing sequences H9 and H7
Mutation of site 36 in sequence H9
Coalescing sequences H9 and H2
Mutation of site 16 in sequence H9
Mutation of site 19 in sequence H9
Mutation of site 23 in sequence H9
Mutation of site 24 in sequence H9
Coalescing sequences H9 and H16
Mutation of site 27 in sequence H9
Coalescing sequences H9 and H15
Mutation of site 13 in sequence H9
Coalescing sequences H1 and H9

*nad*6

```
Position             112222333344555
                    234176889235649015
                    381666183620934357

Site Number                  111111111
                    123456789012345678

Consensus           001000000001010000
Character Type      --i------i-i-ii-ii
H1          (  3)  ..0........0.0....
H2          (  3)  1............11.1
H3          (  2)  .1...1...1........
H4          (  1)  .......1.1......1.
H5          (  2)  ...1..........1.11
H6          (  2)  ....1.1.1.....1.11
H7          (  6)  .........1......1.
H8          (  2)  ...........0.0....
H9          (  4)  ..0.......1.10....
H10         (  2)  ..0..........0....
```

Mutation of site 1 in sequence H2
Mutation of site 2 in sequence H3
Mutation of site 4 in sequence H5
Mutation of site 5 in sequence H6
Mutation of site 6 in sequence H3

Mutation of site 7 in sequence H6
Mutation of site 8 in sequence H4
Mutation of site 9 in sequence H6
Mutation of site 11 in sequence H9
Mutation of site 13 in sequence H9
Mutation of site 16 in sequence H2
Coalescing sequences H5 and H6
Coalescing sequences H4 and H7
Coalescing sequences H9 and H10
Recombination in sequence H4 after site 16; prefix is new sequence 11
Coalescing sequences H3 and 11
Mutation of site 10 in sequence H3
Recombination in sequence H4 after site 17; prefix is new sequence 12
Coalescing sequences H5 and 12
Mutation of site 17 in sequence H5
Coalescing sequences H1 and H4
Coalescing sequences H2 and H5
Mutation of site 15 in sequence H2
Mutation of site 18 in sequence H2
Coalescing sequences H2 and H3
Mutation of site 14 in sequence H2
Recombination in sequence H2 after site 11; prefix is new sequence 13
Coalescing sequences H8 and 13
Mutation of site 3 in sequence H8
Coalescing sequences H1 and H8
Coalescing sequences H2 and H9
Mutation of site 12 in sequence H2
Coalescing sequences H1 and H2

cox2

```
Position           1112234455
                  60467986700
                  66775394003

Site Number                11
                  12345678901

Consensus         00010000100
Character Type    ----ii--ii-
H1          ( 10) ...0....0..
H2          (  2) 11......0..
H3          (  2) ..1.111..1.
H4          (  2) ..........1
H5          (  6) ....11.1.1.
H6          (  2) ..........
```

Mutation of site 1 in sequence H2
Mutation of site 2 in sequence H2
Mutation of site 3 in sequence H3
Mutation of site 7 in sequence H3
Mutation of site 8 in sequence H5
Mutation of site 11 in sequence H4
Coalescing sequences H3 and H5

Coalescing sequences H4 and H6
Mutation of site 5 in sequence H3
Mutation of site 6 in sequence H3
Mutation of site 10 in sequence H3
Coalescing sequences H3 and H4
Mutation of site 9 in sequence H3
Coalescing sequences H2 and H3
Mutation of site 4 in sequence H2
Coalescing sequences H1 and H2

Table S5. Sites under negative selection identified with the Single Likelihood Ancestor Counting (SLAC) method, the Fixed Effect Likelihood (FEL) method, and the Random Effect Likelihood (REL) method from Datamonkey (http://www.datamonkey.org/) among the mitochondrial genes used in this study.

Codon	SLAC dN-dS	SLAC p-value	FEL dN-dS	FEL p-value	REL dN-dS	REL Bayes Factor
***atp*6**						
5	-2.02	0.004	0	not significant	-0.864	not significant
13	-6.287	0.002	0	not significant	-0.924	not significant
17	-1312.74	not significant	-2478.55	0.007	-4.979	3093770000
49	-941.727	not significant	-823.572	0.07	-4.978	16299.9
76	-1005.16	not significant	-495.579	not significant	-4.979	35664.6
89	-656.369	not significant	-848.593	0.067	-4.98	123228
114	-656.369	not significant	-637.613	0.079	-4.983	296833
135	-941.727	not significant	-5604.07	0.017	-4.983	61813.4
159	-371.719	not significant	-854.531	not significant	-1.886	519.247
160	-1704.28	not significant	-752.215	0.076	-4.977	15332.2
192	-941.727	not significant	-4616.28	0.019	-4.975	50506.5
202	-26.136	0.009	0	not significant	-0.928	not significant
219	-13.536	0.008	0	not significant	-0.682	not significant
***cox*2**						
41	-62.682	0.02	-2818.6	0.001	-13.023	not significant
49	-17.056	not significant	-655.078	0.042	-13.023	not significant
81	-0.711	0.017	0	not significant	-0.275	not significant
85	-62.618	0.024	-905.113	0.021	-13.023	not significant
91	-34.112	not significant	-1926.2	0.003	-13.023	not significant
133	-0.711	0.017	0	not significant	-0.363	not significant
136	-0.36	0.022	0	not significant	-0.656	not significant
142	-0.711	0.017	0	not significant	-0.275	not significant
148	-20.873	not significant	-1097.12	0.049	-13.023	not significant
150	-41.745	0.074	-590.172	0.025	-13.023	not significant
***nad*3**						

18	-49.405	not significant	-325.08	not significant	-3.515	93702000
23	-43.938	not significant	-905.179	0.036	-3.515	20732.3
39	-41.551	not significant	-774.534	0.079	-3.515	19221.7
66	-43.938	not significant	-994.362	0.066	-3.515	20597.6
78	-43.938	not significant	-573.646	0.077	-3.515	18235.1
95	-43.938	not significant	-742.292	0.059	-3.515	19776.9
nad5						
28	-77.559	not significant	-704.592	0.055	-4.087	14380.3
51	-91.365	not significant	-500.55	0.069	-4.088	12688.6
74	-77.559	not significant	-597.771	0.051	-4.09	13662.6
78	-91.365	not significant	-1276.65	0.056	-4.084	15251
83	-155.118	not significant	-381.726	0.055	-4.088	81985400
98	-91.365	not significant	-836.899	0.062	-4.085	13666.6
139	-182.73	0.089	-2206.13	0.01	-4.084	143164000
141	-77.559	not significant	-498.895	0.061	-4.09	13212.1
143	-91.365	not significant	-241.371	not significant	-4.085	9474.17
200	-77.559	not significant	-188.748	not significant	-4.09	8314.29
259	-155.118	not significant	-404.167	0.051	-4.088	85656100
287	-232.676	0.037	-588.98	0.014	-4.089	8.38546E+11
294	-77.559	not significant	-675.131	0.041	-4.091	14305.3
312	-77.559	not significant	-597.771	0.045	-4.091	13925.2
314	-77.559	not significant	-198.192	not significant	-4.089	8616.07
317	-77.559	not significant	-360.392	not significant	-4.087	11857.5
326	-155.118	not significant	-428.555	0.046	-4.088	89614800
353	-91.365	not significant	-500.55	0.078	-4.087	12674.4
356	-182.73	0.089	-407.101	0.089	-4.084	86588600
361	-91.365	not significant	-484.34	0.079	-4.087	12595.3
372	-116.769	not significant	-3201.65	0.024	-4.087	15751.4
392	-77.559	not significant	-700.426	0.041	-4.091	14213.4
410	-77.559	not significant	-173.26	not significant	-4.087	7721.2

416	-91.365	not significant	-236.045	not significant	-4.087	9352.79
439	-77.559	not significant	-611.119	0.064	-4.087	13765.8
469	-77.559	not significant	-167.337	not significant	-4.09	7519.01
472	-91.365	not significant	-491.837	not significant	-4.083	13078.1
486	-77.559	not significant	-489.08	0.074	-4.088	13160.6
513	-77.559	not significant	-668.16	0.058	-4.087	14261.6
521	-77.559	not significant	-597.77	0.051	-4.09	13662.6
527	-117.367	not significant	-4027.33	0.007	-4.087	18364
***nad*6**						
7	-78.858	not significant	-1096.19	0.017	-1.828	36584.1
12	-45.486	not significant	-327.101	0.085	-1.828	31925.3
13	-78.858	not significant	-88.363	0.157	-1.828	21615.4
38	-63.951	not significant	-339.001	0.04	-1.828	33849.7
58	-67.021	not significant	-240.074	0.088	-1.828	30962
88	-63.951	not significant	-262.245	0.053	-1.828	31509.2
93	-73.924	not significant	-303.302	0.073	-1.828	32033.5
97	-63.951	not significant	-273.948	0.035	-1.828	31566.4
99	-0.182	0.008	0	not significant	-0.954	not significant
108	-77.595	not significant	-311.117	0.061	-1.828	36626.6
110	-63.951	not significant	-65.223	not significant	-1.828	17512.2
116	-157.715	0.073	-1171.84	0.002	-1.828	342561000
123	-36.255	not significant	-77.983	not significant	-1.828	19807.3
147	-63.951	not significant	-76.223	not significant	-1.828	19666.9
164	-63.951	not significant	-307.78	0.045	-1.828	31904.1
167	-63.951	not significant	-286.269	0.048	-1.828	31578
169	-127.902	not significant	-188.614	0.032	-1.828	218353000
171	-127.902	not significant	-187.33	0.025	-1.828	226639000
177	-85.22	not significant	-22082	0.008	-1.828	35926
185	-63.951	not significant	-307.013	0.032	-1.828	31946.5