Molecular Evolution and Mutation Accumulation Lines in the Nematode *Pristionchus pacificus*

D i s s e r t a t i o n

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Ruxandra Molnar
aus Bukarest
Rümanien

Tübingen
2012

Tag der mündlichen Qualifikation:	1.08.2012
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Ralf J. Sommer
2. Berichterstatter:	Prof. Dr. Gerd Jürgens

Publications

Teile der Arbeit wurden bereits veröffentlicht oder sind zur Veröffentlichung eingereicht:

1. **Ruxandra I. Molnar**, Gabi Bartelmes, Iris Dinkelacker, Hanh Witte, and Ralf J. Sommer:

   Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*

   Molecular Biology and Evolution 2011, 28(8):2317–2326.

2. **Ruxandra I. Molnar**, Hanh Witte, Iris Dinkelacker, Laure Villate, and Ralf J. Sommer:

   Tandem Repeat Patterns and Mutation Rates in Microsatellites of the Nematode Model Organism *Pristionchus pacificus*

   G3 (Bethesda) 2012, 2(9):1027–1034.

Acknowledgments

This work was undertaken in the Department of Evolutionary Biology, at Max Planck Institute for Developmental Biology in Tübingen, Germany, under the guidance of Prof. Dr. Ralf J. Sommer.

I would like to sincerely thank my supervisor **Ralf J. Sommer**, for the encouragement, six years of patience, and trust he has given me throughout my thesis work. I have learnt about the worm from you, all about its ecology and evolution. Thank you for supporting me and, at times, criticising me.

I am grateful to **Prof. Dr. Gerd Jürgens** from the Department of Developmental Genetics at the University of Tübingen, and from the Department of Cell Biology at the Max Planck Institute, for agreeing to be the examiner of my thesis.

Dr. Richard Neher and **Prof. Dr. Rolf Reuter**, I appreciate you are taking part as examiners at my thesis defense.

Matthias, **Metta**, and **Melanie**, thank you for translating my summary into German.

I have spent many years in this lab... During the wet-lab work, I have stayed in Lab-4. I would like to thank **Hans**, **Akira**, **Nadine**, **Lisa**, **Simone**, **Robbie**, **Iris**, **Heike**, with whom I have shared the lab at times, for providing nice atmosphere and nice discussions, and for coping with my messiness. If you miss any pen or pencil.. you might consider asking me where it is. **Hans**, I will never forget your experiments: Do worms like coffee, or spicy sauce? or, How long do they survive extreme dessication on agar plates? **Akira** and **Hans**, I thank you for the developmental biology lessons. **Federico** joined later on and brought some south american "air" and the nictation experiments done in collaboration with Isabella certainly made fun for all of us.

When I finished the wet-lab work, I moved to the bioinformatics room. There, **Amit**, **Andreas**, and **Christian** cheered up the mood with nerdy discussions. I miss them! And next time we work in the same lab, don't give me space 'cause I take it! **Christian**, thanks for the Perl lessons that made my life way easier while analysing microsatellites. **Amit**, you always had food for thought and food for stomach. Thanks for all the sandwiches. **Andreas**, thanks for the zoology and, now, genomics discussions and thanks for calming me down, at times. Give a hug to Suska and Fio from me.

All those gels I've been pouring... Lab-1 might have experience an overload. And also, some of the wet-lab work happened in that room. **Gabi**, **Isabella**, **Erik**, **Manuela**, **Vahan**, **Gilberto**, **Huiyu** thank you for not throwing me out of the room and for all the long conversations. **Gabi**, thanks for all your help for the mitochondria paper.

I also want to thank to **Dan (with Barbara and Melinda)**, **Matthias**, **Ben**, **Gilberto**, **Stephan**, **Ray**, **Dave**, **Katy**, **Angela**, **Melanie**, **Min**, for being good labmates and for various encouragements they gave me over the years. The lab runs smoothly due to our great technical team, **Gabi**, **Hanh**, **Heike**, **Metta**, **Iris**. Thank you for endless supply of PCR enzyme, BigDye, and patience. Thanks to **Tamara**, who prepared all those agar plates I have used during the experiments and to **Gertrud**, who helped into keeping clean the lab and for supplying all the yellow-tip boxes, to **Christa** and **Andrea** for giving me off-hours access to the sequencing machine and for processing the tone of plates I have brought in six years.

A special kind of thank you to **Werner**. Your phylogeny and molecular biology help was invaluable to me.

I would like to thank the Streit Group, especially **Adrian**, for all the great ideas during the lab meetings, **Alex** for the Romania-Germany transition, and **Olga** for the last two years of "chatting" and comparisons between Romania and Russia.

Outside the lab the life was different. I am grateful to **Beth** and **George**, and **Laure** and **Virgil** for all the spontaneous visits and long evenings in great company. Thank you **Isabella** and **Simone** for being my friends and for showing me the bright side of a bad day. Thank you **Marco** and **Yu-Hua**, my favourite flatmates in the whole world.

My dear **Borik**, thank you very much for being next to me and supporting me. Thank you for all those breathing lessons at over 100 m; and for the warm meals at late hours...

Nothing would have been possible without **Mama**, **Tata**, and **Ioana**, and my greatparents. To you I owe everything I am. Thank you for encouraging me, for believing I can do it.

Contents

List of Abbreviations

atp6 and *atp8*	genes for ATP synthase subunits 6 and 8
cox1–cox3	genes for cytochrome oxidase *c* subunits 1–3
cytb	gene for cytochrome oxidase *b*
nad1–6 and *nad4L*	genes for NADH dehydrogenase subunits 1–6 and 4L
rrnS and *rrnL*	genes for small and large mitochondrial RNA subunits
A, C, T, G, U	adenine, cytosine, thymine, guanine, uracil
ATP	adenosine-5'-triphosphate
bp, kb, Mb	base pairs, kilobase pairs, megabase pairs
DNA	deoxyribonucleic acid
EGF	epidermal growth factor
IAM	Infinite Alleles Model
MA	mutation accumulation
MP	maximum parsimony
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
nDNA	nuclear DNA
NJ	neighbour-joining
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
R, Y	purine, pyrimidine

RNA	ribonucleic acid
SMM	Stepwise Mutation Model
SR-DNA	simple repetitive DNA
SSLP	simple sequence length polymorphisms
SSR	simple sequence repeats
SSU	small-subunit rDNA
STR	short tandem repeats
TRF	Tandem repeats finder
tRNA	transfer RNA
VNTR	variable number of tandem repeats

List of Figures

List of Tables

Summary

The nematode *Pristionchus pacificus* has been established as a model system for modern evolutionary studies. Evolutionary reconstruction of the natural history of organisms requires knowledge about the development, ecology, and phylogeny of species. Mutations are the source of natural variation, hence studies of mutational processes improve the understanding of the natural history of an organism. Mutation accumulation (MA) lines experiments facilitate the study of spontaneous mutation rates over many generations and offer the possibility of inferring calibrated phylogenies for taxa that lack fossil records. In this study, we used the MA lines setup to calculate mutation rates at the level of the mitochondrial genome and microsatellite loci of the nuclear genome of *P. pacificus*.

Mitochondrial DNA (mtDNA) evolves rapidly in populations, is usually transmitted maternally without intermolecular recombination and has therefore been intensively used for phylogeographic studies. The *P. pacificus* mtDNA is $15,955$ bp in length and contains all the known mitochondrial genes. In the 142^{nd} generation of the 82 MA lines, we found an overall mutation rate of 7.6×10^{-8} per site per generation. The unusual presence of a suppressor transfer RNA for the codon UAA has, most likely, influenced the spectrum of observable mutations, in that the lines containing such a premature STOP codon were actually able to survive until the end of the experiment. Using the mutation rate estimate in a comparison of the mitochondrial genome of nine *P. pacificus* isolates, we calculated the minimum time to the most recent common ancestor at 10^5–10^6 generations.

Microsatellites are widely used in evolutionary biology, but still little is known about the true nature of their evolution and the factors that affect their frequency, distribution, and mutation. Around $0.59 - 3.83\%$ of the *P. pacificus* genome is composed of di- to hexanucleotides simple tandem repeats. In the MA lines, we analysed a set of 41 randomly chosen microsatellites, in order to evaluate the mutation rates at various loci in the nuclear genome. A total of 31 mutations yielded a mean mutation rate of 7.4×10^{-5} per allele per generation. We observed a strong correlation between allele size and mutation rate, a tendency towards lengthening, and little evidence of multistep mutations. The mutation rates obtained for individual markers provide a powerful tool for divergence time estimates in future studies.

These studies provide first insights into mutation processes that shape the genome architecture of *P. pacificus* and help further evolutionary studies of this organism.

Zusammenfasung

Der Nematode *Pristionchus pacificus* wurde als Modellsystem für evolutionsbiologische Untersuchungen etabliert. Für die Rekonstruktion der naturgeschichtlichen Evolution eines Lebewesens muss man die Entwicklungsbiologie, Ökologie und Phylogenie der Art kennen. Mutationen sind die Grundlage der natürlichen Variation und deshalb verbessern Untersuchungen der Mutationsprozesse unser Verständnis der Entstehungsgeschichte eines Lebewesens. Experimente mit "mutation accumulation lines" ("Mutations-Anhäufungs-Linien", MA-Linien) vereinfachen die Untersuchung spontaner Mutationsraten über viele Generationen und ermöglichen das Erstellen kalibrierter Phylogenien für Taxa ohne Fossilien. In dieser Arbeit verwendeten wir MA-Linien, um Mutationsraten im mitochondrialen Genom und in Mikrosatelliten im nukleären Genom von *Pristionchus pacificus* zu errechnen.

Die mitochondriale DNS einer Population verändert sich schnell und wird normalerweise maternal ohne intermolekulare Rekombination übertragen. Sie wird deshalb häufig für phylogeographische Untersuchungen verwendet. Die mitochondriale DNS von *P. pacificus* ist 15955 bp lang und enthält alle bekannten mitochondrialen Gene. In den 82 MA-Linien fanden wir nach 142 Generationen eine Mutationsrate von $7,6 \times 10^{-8}$ pro Locus und Generation. Ungewöhnlich ist das Vorhandensein einer Suppressor-Transfer RNS für das Codon UAA. Dies beeinflusste höchstwahrscheinlich die Bandbreite der beobachtbaren Mutationen, insofern dass die Linien, die ein solches vorzeitiges STOP-Codon enthalten, tatsächlich bis zum Ende des Experiments überleben konnten. Wir verglichen die Mutationsraten der mitochondrialen Genome von neun *P. pacificus* Stämmen und errechneten, dass der zeitliche Abstand zum letzten gemeinsamen Vorfahren mindestens 10^5–10^6 Generationen beträgt.

Obwohl Mikrosatelliten in der Evolutionsbiologie vielfach verwendet werden, ist dennoch sehr wenig bekannt über ihre Evolution und die Faktoren, die ihre Häufigkeit, Verteilung und Mutationseigenschaften beeinflussen. Zwischen $0,59$ und $3,83\%$ des *P. pacificus* Genoms besteht aus einfachen Di- bis Hexanukleotid Tandem-Wiederholungen. In den MA-Linien untersuchten wir eine Gruppe von 41 zufällig ausgewählten Mikrosatelliten, um die Mutationsraten an verschiedenen Loci im nukleären Genom errechnen zu können. Insgesamt 31 festgestellte Mutationen ergaben eine durchschnittliche Mutationsrate von $7,4 \times 10^{-5}$ pro Allel und Generation. Wir beobachteten einen deutlicher Zusammenhang zwischen Allel-Größe und Mutationsrate, eine Tendenz zur Verlängerung und wenig Hinweise auf das Vorkommen von Mehrschrittmutationen. Die Mutationsraten, die für einzelne Marker errechnet wurden, werden für die Berechnung von Divergenzzeiten in zukünftigen Untersuchungen äußerst nützlich sein.

Diese Arbeit liefert erste Einblicke in Mutationsvorgänge, die die genomische Architektur von *P. pacificus* gestalten. Außerdem können die in dieser Arbeit errechneten Mutationsraten für verschiedene andere evolutionsbiologische Untersuchungen in diesem Tier verwendet werden.

Part I

Background

Chapter 1

The Nematode *Pristionchus pacificus*

a well-rounded nematode – Ray L. Hong

This chapter will introduce the nematode *Pristionchus pacificus*. This free-living nematode has been sampled in a flower garden in Pasadena in 1988 and described for the first time by Sommer et al. [1996], as "useful for genetic, developmental and molecular biological studies". Comparative studies between *P. pacificus* and the nematode model organism *Caenorhabditis elegans* have been initiated with the construction of a cDNA library, and the cloning of a homologue of the *C. elegans let-60 ras* gene.

One of the best understood processes in animal developmental biology is the organogenesis of the *C. elegans* vulva [Wang and Sternberg 2001]. Despite similar morphology of this organ amongst the different nematodes, many changes are displayed at the molecular level. They range from the induction mechanism [Sommer 2000], to the number of cells that will eventually form the vulva [Félix et al. 2000], and to the evolution of new organizing centers [Jungblut and Sommer 2000]. The developmental studies on *P. pacificus* and other nematodes have been coupled with robust phylogenetic framework [Kiontke et al. 2007]. Phylogenetic efforts are undertaken for more than 20 species of *Pristionchus* and more than 400 strains of *P. pacificus* in order to understand the ecology of these nematodes [Herrmann et al. 2006a,b, 2007; Mayer et al. 2007; Herrmann et al. 2010; Morgan et al. 2012].

1.1 Evo-devo Studies

«Thus, *P. pacificus* sp. n. can serve as a "satellite" organism for comparison to the model system *C. elegans*. Our longterm goal is a detailed understanding of vulva development in *P. pacificus* sp. n.»

Ralf J. Sommer in *Morphological, genetic and molecular description of Pristionchus pacificus sp. n. (Nematoda: Neodiplogastridae)*, Fundamental and Applied Nematology 1996.

"I cannot really describe how triumphant I felt working with E1, backcrossing it to wild-type N2 with the male cultures I had previously established, and proving that the resulting heterozygotes segregated it in the classic Mendelian ratio of $1:3$. Getting a mutant of a complex organism and confirming Gregor Mendel in only two weeks was most satisfying" [Brenner 2009, p. 413]. Key features of the nematode *Caenorhabditis elegans* are the ease of culture, growth, observation, analysis, and genetic manipulation. *C. elegans* is about 1 mm in length at adulthood, with a life cycle of only three to four days when grown on agar plates covered with *Escherichia coli* at $20\,^{\circ}\mathrm{C}$. It is a self-fertilizing hermaphrodite, with possibility of genetic crossing due to fertile males that appear in culture, under certain conditions [Brenner 1974]. *C. elegans* nematodes can be stored in liquid nitrogen [Sulston and Brenner 1974], they are transparent throughout the life cycle, and they have fixed cell-lineage [Chalfie et al. 1981]. These characteristics made *C. elegans* an ideal nematode model organism, and alongside *Drosophila melanogaster*, it is, now, the subject of research for more than 50 years.

Studies in *C. elegans* have been invaluable for understanding a wide range of biological phenomena, but generalizations from one species to a phylum are not always valid, not to mention, the ones made to the entire animal kingdom [Hong and Sommer 2006]. Therefore, an expansion of the *C. elegans* model was essential. The nematode *Pristionchus pacificus* has been established as a satellite model organism to *C. elegans*, after it has been initially surveyed for vulva patterning [Sommer and Sternberg 1996]. Evolutionary-developmental biology (evo-devo) investigates the evolution of developmental processes, aiming for a mechanistic understanding of phenotypic variation [Sommer 2009]. Evo-devo studies rely on model species for which the genetic methodology is well established. Forward genetics, reverse genetics and transgenesis methods are a must, while genome projects, phylogenetic reconstructions and genome-wide association studies make possible the investigation of genome architecture and directionality of evolutionary changes [Sommer 2009].

Similar to *C. elegans*, *P. pacificus* is a self-fertilizing hermaphrodite with occasional spontaneous males [Sommer et al. 1996], it can be raised on OP50 *E. coli* bacterial cultures

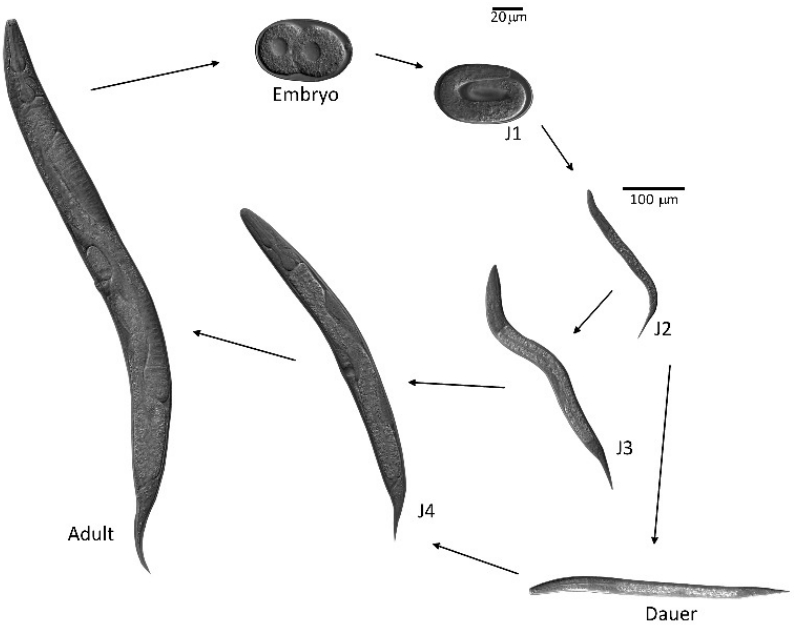


Figure 1.1 – Life cycle of *Pristionchus pacificus* at $20\,^{\circ}\mathrm{C}$. *P. pacificus* propagates through four juvenile stages, called J1 to J4. The J1 to J2 molt is embryonic and only the J2 stage hatches from the egg. Dauer formation occurs as the alternative J3 stage. The direct life cycle takes approximately 4 days at $20\,^{\circ}\mathrm{C}$. Pictures courtesy of Metta Riebesell.

on agar plates, it has a short generation time at $20\,^{\circ}\mathrm{C}$ in standard laboratory conditions, a large brood size, and it can be cryopreserved. It is amenable to standard techniques, such as cell lineaging [Sommer 2001], cell ablations [e.g., Jungblut and Sommer 2000], mutagenesis (EMS, ENU, UV) [e.g., Kenning et al. 2004; Gutierrez and Sommer 2007], deletion-library screening [e.g., Tian et al. 2008], morpholino transformation [Pires-daSilva and Sommer 2004], and transgenesis [Schlager et al. 2009].

Two developmental processes have been studied in great genetic and molecular details: vulva formation [Eizinger and Sommer 1997; Jungblut and Sommer 2000; Schlager et al. 2006; Zauner and Sommer 2007, and others] and dauer development [Ogawa et al. 2009, 2011; Sinha et al. 2012]. It has been revealed that, although these two processes are conserved both in *P. pacificus* and in *C. elegans*, different molecular regulations are acting to ensure them. The regulation of vulva development in *P. pacificus* is achieved via the Wnt pathway from two signalling centers – the somatic gonad and the posterior body region [Zheng et al. 2005; Tian et al. 2008; Wang and Sommer 2011]. This is fundamentally different from *C. elegans*, where vulva formation requires the coordination of three developmental pathways: EGF, Notch, and Wnt [Eisenmann et al. 1998; Myers and Greenwald 2007; Green et al. 2008]. Recent studies have shown that two transcription

factors crucial for the dauer formation, DAF-12 and DAF-16, are well conserved between *C. elegans* and *P. pacificus* [Ogawa et al. 2009, 2011], but different downstream targets are activated to complete the process [Sinha et al. 2012]. Wnt pathway and the regulatory pathway that includes DAF-12 have been coopted independently for gonadogenesis [Rudel et al. 2008], and, respectively, mouth form morphology [Bento et al. 2010].

The life cycle of *P. pacificus* is composed of four larval stages, J1 to J4 (Figure 1.1). The first-stage larva (J1) is retained in the egg, and the second-stage larva (J2) is the stage that hatches, a common feature of all Diplogastrid nematodes. In response to environmental cues indicating starvation or overcrowding, the *P. pacificus* J2 develops into a dauer larva instead of a third-stage larva (J3). Dauer larvae are key components of the survival and dispersal strategy of nematodes. Their development is arrested, they do not feed because their mouths are plugged as a result of retaining the cuticle from the J2 stage, and they are resistant to many abiotic factors [Riddle and Albert 1997]. *Pristionchus* dauer larvae can survive for up to one year in laboratory conditions [Mayer and Sommer 2011]. Dauer formation and recovery from the dauer stage are tightly regulated in order to balance the utilisation of local food resources with the needs of an expanding population to disperse and forage for food. There are morphological similarities between dauer larvae of free-living nematodes and infective larvae of parasitic nematodes, and this lead to the idea that dauer larvae represent, in fact, a pre-adaptation towards the parasitic lifestyle [Sudhaus 2010]. Also, it has been shown that the same endocrine module is responsible for both dauer and infective larvae formation [Ogawa et al. 2009].

1.2 Phylogeny and Ecology

«What is true for the human species, that no two individuals are alike, is equally true for all other species of animals and plants.»

Ernst Mayr (Evolutionary biologist (1904–2005)), *Evolution and Anthropology*, 1959.

The evo-devo differences between *P. pacificus* and *C. elegans* might be due to distinct lifestyles. Comparative studies, like the ones mentioned in the previous section, require a robust phylogeny. Surveys of terrestrial and marine fauna bring the estimate of the number of nematode species at 100 million, making the Phylum Nematoda one of the most diverse in the animal kingdom, with numerous lifestyles, from free-living species in soil, freshwater or ocean sediments, to parasitic species in humans, livestock and plants [Holterman et al. 2006].

There are several influential studies that aimed at characterizing the relationships

between individuals at the level of the entire Phylum Nematoda. Based on the small subunit (SSU) sequences from 53 taxa, Blaxter et al. [1998] defined five major clades, named Clade I to Clade V. Following this nomenclature, the order Diplogasterida, to which *P. pacificus* belongs, is part of Clade V, and it is sister group to the order Rhabditida, which is paraphyletic (Figure 1.2). More precisely, Diplogasterida is in the same clade as the rhabditid family Rhabditoidea, which includes *C. elegans* [Blaxter et al. 1998].

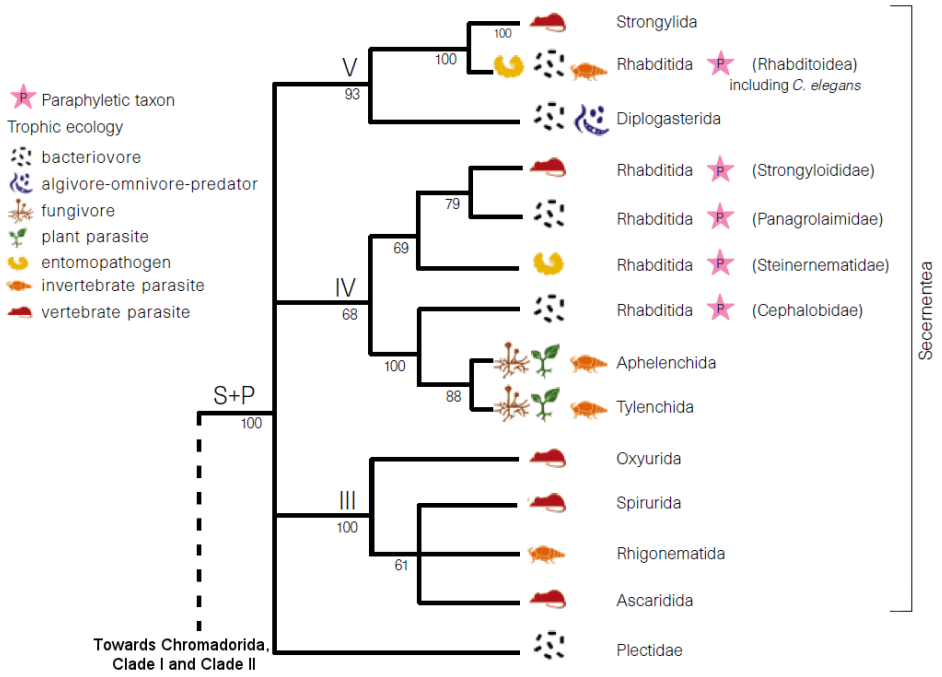


Figure 1.2 – A subset of the phylogenetic hypothesis for the Phylum Nematoda according to Blaxter et al. [1998] Dendogram summarizing the results of MP and NJ analysis. Numbers at the branch point represent the maximal percentage bootstrap support. Five major clades have been defined, but here Clades I and II are not shown. S + P indicates the clade Secernentea plus Plectidae. Trophic ecologies are indicated. Accepted taxa found to be paraphyletic are indicated by a starred P. This figure is modified from Blaxter et al. [1998]

Holterman et al. [2006] performed the phylogenetic reconstructions of 339 nematode taxa, inferred from nearly full-length SSU rDNA sequences. In their reconstruction, the Phylum Nematoda has been divided into 12 major groups, named Clade 1 to Clade 12. According to this nomenclature, the family Diplogasteridae, which includes *P. pacificus*, is part of Clade 9 together with families like Rhabditidae and Neodiplogasteridae. Clade 9 *sensu* Holterman et al. [2006] corresponds entirely to Clade V *sensu* Blaxter et al. [1998].

In the most recent attempt to characterise the Phylum Nematoda, van Megen et al. [2009] based their phylogeny on 1,215 full-length SSU rDNA sequences from representative nematodes. In this survey, they inferred the same tree topology, as in Holterman et al.

[2006], but they have proposed further subdivisions for some of the clades, in order to ease up the discussion. Accordingly, the Diplogasteridae family became part of Clade 9A.

The family Diplogasteridae includes more than 300 species of free-living or insect-associated nematodes, grouped in 28 genera [Sudhaus and Fürst von Lieven 2003]. The most recent study of rhabditid nematodes, conducted by Kiontke et al. [2007], has placed the diplogastrids as sister clade to Eurhabditis, and implicitly *Caenorhabditis* (Figure 1.3), which is consistent with Blaxter et al. [1998], but it is in contradiction with Holterman et al. [2006].

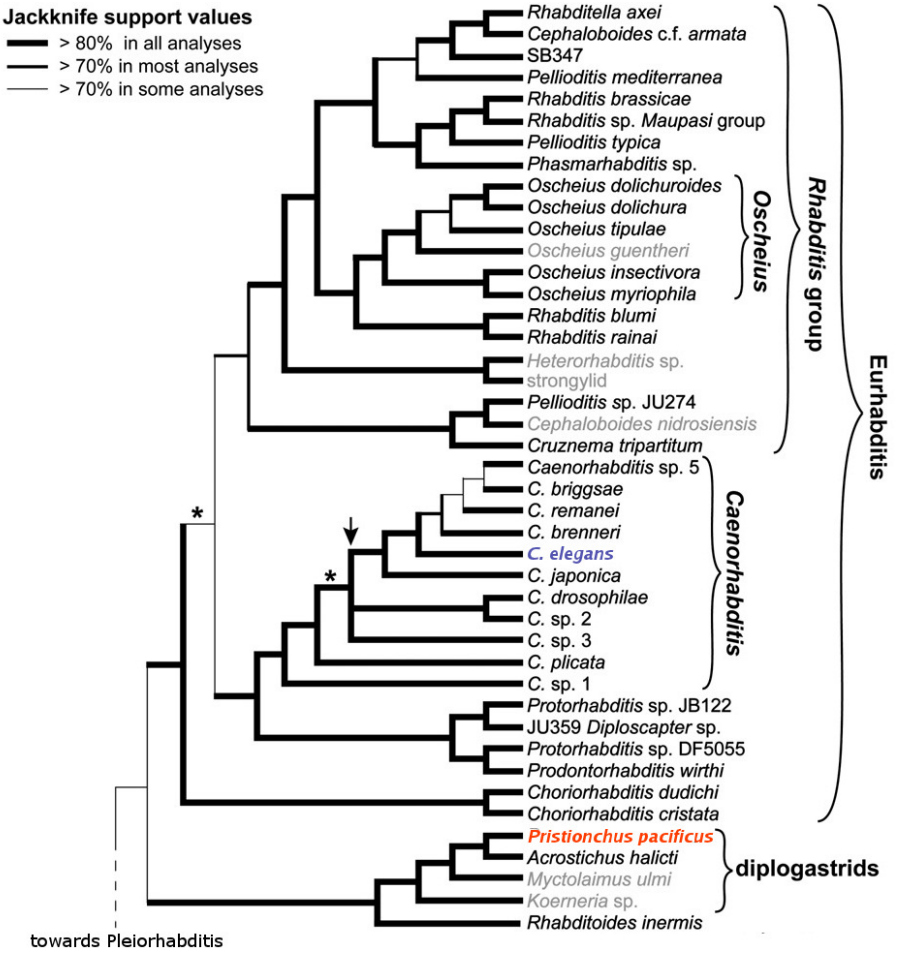


Figure 1.3 – Part of the rhabditid phylogeny as shown in Kiontke et al. [2007]. A hypothesis for the phylogenetic relationships of rhabditids as derived from six weighted MP jackknife analyses performed with concatenated sequences of genes for SSU and LSU rRNA and the largest subunit of RNA polymerase II. Thickness of branches denote the level of jackknife support. Taxa in gray were absent from some of the analyses; the Bayesian analyses recovered the same tree with the exception of braches marked with asterisks (*).

The relationship of the *Pristionchus* genus with other diplogastrids has been resolved, based on SSU sequences and ribosomal protein genes, with the use of a series of phylogen-

tic inferences (maximum parsimony (MP), neighbour-joining (NJ), maximum likelihood (ML), and bayesian inference (BI)) [Mayer et al. 2009] (Figure 1.4). The phylogenetic position of *P. pacificus* within the *Pristionchus* genus has been established in a similar study [Mayer et al. 2007]. This phylogeny serves as a framework for mechanistic studies in developmental biology, evolution, behaviour and ecology [Mayer et al. 2007]. Systematic field studies in North and South America, Asia, Africa and Europe had as outcome the identification of more than 20 *Pristionchus* species, both hermaphroditic and gonochoristic [Herrmann et al. 2006a,b, 2007; Mayer et al. 2007]. Amongst these, *P. pacificus* is the most widely distributed species of the genus [Zauner et al. 2007].

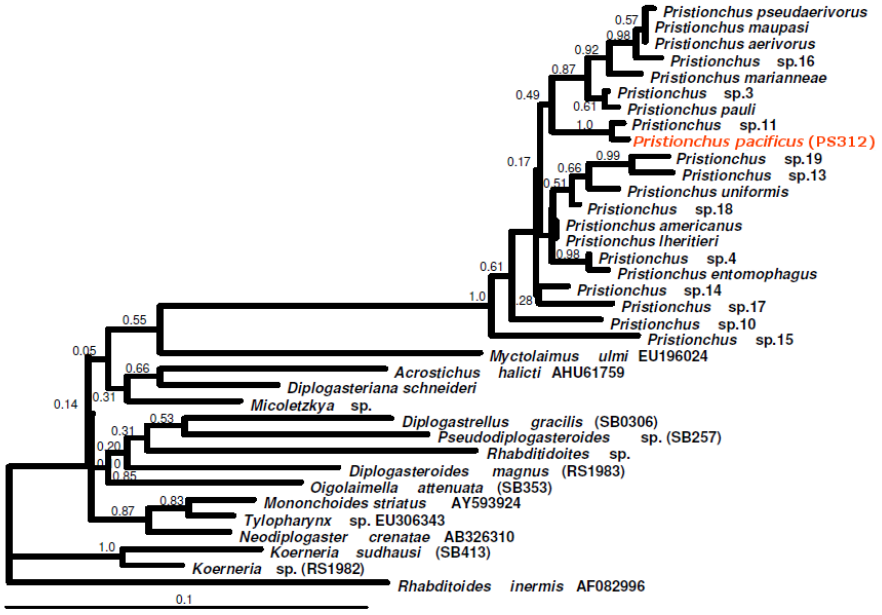


Figure 1.4 – Maximum likelihood tree of diplogastrid *SSU* sequences. Tree as it was presented in Mayer et al. [2009]. Values indicate bootstrap support; strain designations are shown in parentheses.

The natural habitat of the model organism *C. elegans* is still under debate. Most of the *C. elegans* isolates stem from composts heaps, where the species is found, sometimes in association with various invertebrates such as millipedes, isopods, insects, snails and slugs [Barrière and Félix 2005]. Contrary to this, *Pristionchus* species are found in a necromenic association with beetles. Particularly, *P. pacificus* is found on scarab beetles [Herrmann et al. 2007]. Nematodes in the dauer stage stay associated with the beetles until the death of their host [Weller et al. 2010], after which, their development is resumed and the nematodes feed on the microbes on the beetle carcasses [Herrmann et al. 2006a]. Several studies showed an almost species-specific association between the different *Pristionchus* species and their hosts [Herrmann et al. 2006a,b, 2007], as the

nematodes are attracted to sex-pheromone of insects related to their own ecology [Hong et al. 2008].

More than 400 strains of worldwide origin are available to date, for various laboratory investigations. In the study by Herrmann et al. [2010], 160 of them have been phylogenetically characterized based on a *nad6-nad4L* fragment of the mtDNA (Figure 1.5). Based on branch length and support values, the tree is split in four distinct clades, named Clade A to Clade D [Herrmann et al. 2010]. Clade A consists of strains originating in Asia, USA, Bolivia and La Réunion Island. Clade C contains most of the strains sampled on La Réunion and only few isolates from USA, Montenegro, and Madagascar. Samples from South Africa, Bolivia, La Réunion and Switzerland group together and form clade D [Herrmann et al. 2010]. Subsequent samplings added several more strains from La Réunion in Clade B, which is composed of strains from this region only [Mayer WE, personal communication].

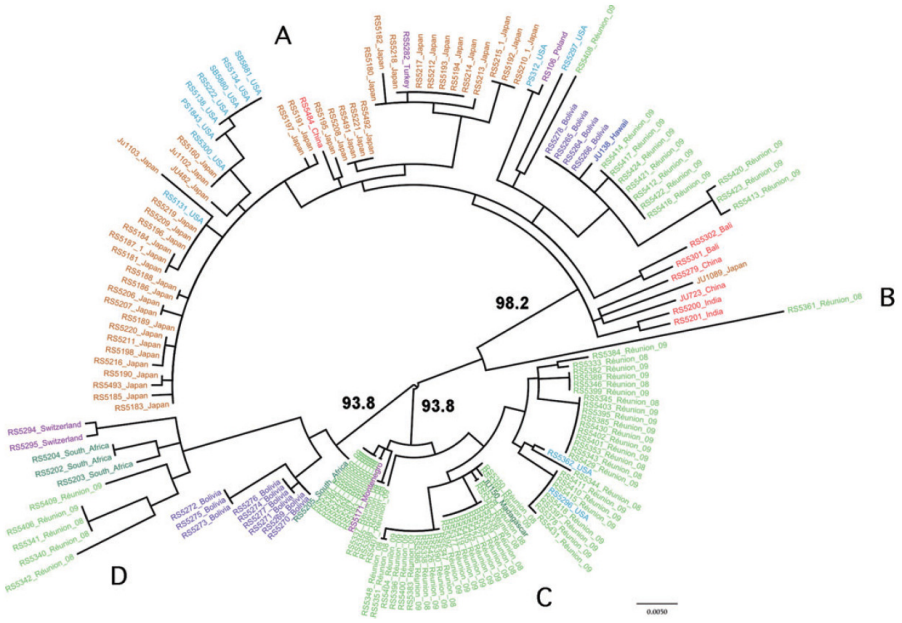


Figure 1.5 – Phylogenetic relationships of *Pristionchus pacificus* strains. Radial representation of an unrooted ML tree, inferred from mitochondrial *nad6-nad4L* fragment. Taxons are labelled by strain number, geographic origin, and sampling year. Taxon colours indicate the geographic origins of the *Pristionchus* strains. Values on the branches indicate NJ bootstrap support vales for the major clades A, C, and D. This figure is modified from Herrmann et al. [2010].

Two regions have been intensively studied with respect to *P. pacificus*: Japan [Herrmann et al. 2007] and La Réunion Island in the Indian Ocean [Herrmann et al. 2010; Morgan et al. 2012]. In Japan, the oriental beetle *Exomala orientalis* has been sampled at more than 60 sites, and, although their infestation rate was rather low, still 40 *P. pacificus* strains

have been isolated [Herrmann et al. 2007]. In the same study, the authors have shown that *E. orientalis* is also an important host species for *P. pacificus* nematodes in North America. It was proposed that the nematodes were transported together with the beetle, when this was introduced from Japan to the East Coast of the United States. Chemotaxis assays showed that *P. pacificus* was the only species attracted to the *E. orientalis* sex-pheromone. This was the first time that a specific beetle host was identified for *P. pacificus* [Herrmann et al. 2007]. All the samples from Japan cluster in Clade A, showing low nucleotide diversity. The samples from La Réunion, on the other hand, are present in all four clades of the tree, separated by quite long branches. La Réunion Island is only 2–3 million years old. Nevertheless the nematodes isolated have the same high nucleotide diversity found worldwide. This indicates that *P. pacificus* invaded the island several times independently [Herrmann et al. 2007]. Further population genetic studies, focused on multilocus approaches based on microsatellite markers [Morgan et al. 2012], will shed light upon the biogeography of more than 400 *P. pacificus* nematode strains.

1.3 Genomics

«The speedup in sequencing involved using automated reading of DNA sequencing gels through dye-labeling of bases, and breaking down the chromosome into random fragments and sequencing these fragments as rapidly as possible without knowledge of their location in the whole chromosome … Although the same sequence was often obtained up to 10 times, the sequence of the entire chromosome (2×10^9 bp), less a few gaps, was rapidly assembled in the computer over a 9-month period at a cost of about $\$10^6$.»

Bioinformatics: Sequence and Genome Analysis, 2000
(on the first genome assembled – *Hemophilus influenzae* [Fleischmann et al. 1995])

Caenorhabditis elegans was the first metazoan organism to have its complete genome sequenced [The *C. elegans* Sequencing Consortium 1998] and, since then, many more nematodes have been scheduled for sequencing both from terrestrial and marine environments [M Blaxter, personal communication]. The genome of *P. pacificus* has been sequenced in the whole-genome shotgun approach with 10-fold coverage [Dieterich et al. 2008]. Based on previous divergence dates between *C. elegans* and *Caenorhabditis briggsae* [Coghlan and Wolfe 2002] and between *Brugia malayi* and *C. elegans* [Ghedin et al. 2009], it was estimated that *P. pacificus* and *C. elegans* separated 280–430 million years ago.

P. pacificus genome size of 169 Mb is slightly longer than the *C. elegans* genome, and it has 29,000 predicted protein-coding genes, from which 23,500 were confirmed by transcriptomic and proteomic studies [Borchert et al. 2010]. The length increase of the *P. pacificus* genome compared to *C. elegans* is partly due to repeats, and partly due to a

greater number of transcripts [Dieterich et al. 2008]. The repeat coverage is about 17%, and it is subdivided into low-complexity regions, tandem repeats, and complex repeats. From these, the tandem repeats covered aproximatively 3% of the genome. The increased number of transcripts, comparing to *C. elegans*, is considered to be due to the different lifestyle of the two nematodes.

Initially, the analysis of the *P. pacificus* genome was performed alongside other nematodes, *C. elegans*, *C. briggsae*, *B. malayi*, in order to explore the homology relationships of the proteomes, and alongside *D. melanogaster* and *Homo sapiens*, in order to analyse orthologous gene groups. The analysis revealed that around $11,000$ predicted proteins of *P. pacificus* have no counterpart in any of the five organisms included in the comparison. They are called “pioneer genes”, and proteomic analysis confirmed their expression, at least in certain conditions [Borchert et al. 2010]. A set of seven cellulase genes were further analysed by Mayer et al. [2011], and shown to have been acquired from several distinct microbial donors, via horizontal gene transfers (HGT). The presence of such genes has been documented previously for plant parasitic nematodes, like *Globodera* [Smant et al. 1998], *Meloidogyne* [Rosso et al. 1999], or *Bursaphelenchus* [Kikuchi et al. 2005], and, as they show the highest homologies to bacterial equivalents, Yan et al. [1998] proposed that they could have been acquired by HGT. The beetle-associated lifestyle of *P. pacificus* is a perfect source of potential donors for HGT. Analysis based on the codon usage revealed that a substantial fraction of the *P. pacificus*-specific genes originated from insect genomes, via a HGT event, intermediated by viruses as potential carriers [Rödelsperger and Sommer 2011].

1.4 Instead of Conclusion

The various analyses of the genome and ecology of *P. pacificus* are sustained now by sequences of numerous different strains and *Pristionchus* species, that contribute to our understanding of the "well-rounded nematode".

1.4.1 Reason for This Chapter

Why This Is Directly Relevant to My Line of Research

In this chapter, I resumed some of the studies focused on three main research areas for *P. pacificus*: evo-devo, phylogeny, and genomics. In section 1.1, two of the evo-

devo systems most intensively studied in *P. pacificus*, vulva development and dauer formation, are introduced by comparison to *C. elegans*. In the second section, I described the phylogenetic and ecologic framework for *P. pacificus* and related nematodes. In section 1.3, I gave a short overview of the *P. pacificus* genome.

My research focused on two specific parts of the genome, non-coding microsatellite regions and mitochondrial genome. In the next two chapters, I have made a comprehensive introduction upon these two topics.

Chapter 2

Mitochondria

a circle of surprises – David A. Clayton

This chapter will introduce the reader to the world of the mitochondria, a cytoplasmic organelle that harbours its own DNA, the mitochondrial DNA (mtDNA), that, although highly conserved across kingdoms and taxa, it is still highly diverse. We will see that after one century of research, its characteristics are still controversial. The first sections of this chapter will deal with the origins and genetics of the mitochondria and the mtDNA, followed by the organization of the animal mitochondrial genome, with special emphasis on nematodes. At the end, I will discuss the process and reasons of mutation accumulation in the mitochondrial genomes. This finishes up with the original article published in 2011 in the journal *Molecular Biology and Evolution*.

A combination of technical ease of use, and biological and evolutionary properties of clonality, near-neutrality and clock-like nature of its substitution rate had made the mtDNA the most popular marker of molecular diversity [Galtier et al. 2009]. Virtually every molecular study of animal species involves mtDNA haplotyping at some stage. The gene *cox1* has been chosen as the barcode sequence for the animal kingdom [Hebert et al. 2003]. Experimentally, mtDNA has many advantages: it is relatively easy to amplify because it appears in multiple copies in the cell, it has strongly conserved gene content, very few duplications, no introns, and very short intergenic regions [Gissi et al. 2008], its variable regions are flanked by conserved ones, and it is highly variable in natural populations thereby generating signals about population history over relative short time frames. Nevertheless, these reasons for fame don't come without exceptions and there is plenty of room for doubt.

2.1 Origin of Mitochondria

«Each living cell contains in its protoplasm formations which histologists designate by the name of "mitochondria". These organelles are, for me, nothing other than symbiotic bacteria, which I call "symbiotes".»

Paul Portier (French biologist (1866-1962)), *Les Symbiotes*, 1918

One of the major transitions that mark the evolution of complexity is the one from prokaryotes to eukaryotes – the ancestors of mitochondria and chloroplasts, once free-living prokaryotes, became able to replicate only within a host cell [Szathmáry and Maynard Smith 1995]. Other such transitions are from replicating molecules to populations of molecules in compartments, from unlinked replicators to chromosomes, from RNA as gene and enzyme to DNA and proteins, and, consequently, to the "invention" of the genetic code [Szathmáry and Maynard Smith 1995].

Mitochondria (from the Greek *mitos*, thread and *chrondos*, granule) are double-membrane organelles that play a pivotal role in cellular metabolism (production of ATP from oxidative phosphorylation [Saraste 1999], intermediary metabolism and apoptosis), and harbour a small but essential component of an eukaryote's genetic material, which is the mitochondrial DNA. Across all the eukaryotes, the mtDNAs are so different, both in organization and mode of gene expression, that, initially, it was difficult to imagine the nature of their common ancestor, let alone to infer a common bacterial origin. There have been two concurrent theories regarding the origin of mitochondria in eukaryotic cells:

- ⋆ *Filial compartmentalization theory:* introduced by Cavalier-Smith [1975], stipulates that the genomes of organelles have autogenous origins, whereby part of the nuclear genome became incorporated into membrane-enclosed organelles and assumed an almost-independent existence;
- ⋆ *Endosymbiotic theory:* introduced more than a century ago and formulated in its modern form in the 60's by Lynn Margulis [Sagan 1967], argues that the origin of extra nuclear DNA is exogenous. Margulis's insight was that, in the early history of life on earth, some 3.5 billion years ago, when the only living forms were single-cell creatures without complex internal structures, the evolutionary success arose through increasing cellular complexity. This complexity emerged through symbiosis between organisms of different species. The most effective form of symbiosis occurred when the two cell types merged, each contributing to the creation of the eukaryotic cell.

Today, the molecular evidence is overwhelmingly in favour of the endosymbiotic theory.

There are two scenarios that can explain how the symbiosis may have taken place: (a) the mutualism scenario according to which the bacterium is engulfed by the protoeukaryote, and (b) the predation scenario, in which case the by-products of bacteria that survived the regular feeding became useful to the predator [Stearns and Hoekstra 2005]. The complete sequences of the most bacteria-like mitochondrial genome, namely the protozoan *Reclinomonas americana* [Lang et al. 1997], the most mitochondria-like eubacterial genome, *Rickettsia prowazekii* [Andersson et al. 1998], and the mitochondrial sequences from protists [Burger et al. 1996] contributed enormously to our understanding of the origin of all extant mtDNA to a single ancestral protomitochondrial genome.

2.2 Mitochondrial Genetics

«There is nothing new in the conception that there is such a thing as cytoplasmic heredity... What is new is the view that mitochondria carry it.»

E. V. Cowdry (Canadian histologist (1888-1975)), *The Mitochondrial Constituents*, 1918

Mitochondrial DNA has been intensively used, during more than three decades, for phylogeographic [Avise 1998] and conservation genetics studies [Robertson et al. 2007], for tracing ancestry [Cann et al. 1987], for revealing social dynamics [Bass et al. 1996], and for resolving taxonomic uncertainties [Hebert et al. 2003]. It owes its popularity and usefulness as the marker of choice in animal evolutionary studies to some fundamental characteristics, like clonality, neutrality, clock-like evolution, maternal inheritance, and lack of recombination [Dawid and Blackler 1972]. The absence of any mutations spreading through positive selection assures that mtDNA divergence levels roughly reflect divergence times. However, there are matters that complicate these otherwise simple rules and there is a lot of controversy surrounding them, as it will be discussed in the following sections.

2.2.1 Mitochondrial DNA Inheritance

The absence of both paternal mtDNA and heterologous recombination are hallmarks of mtDNA inheritance. The nuclear genome generally transmits copies of itself to the next generation via fair meiosis involving strict segregation. In contrast, cytoplasmic genome segregation is less equitable and more complex [Barr et al. 2005]. A typical somatic cell contains $500–1000$ mitochondria and an oocyte contains some $10^4–10^5$ mitochondria, each with a few DNA molecules [Satoh and Kuroiwa 1991]. Two to ten

of these mtDNA molecules may aggregate into nucleoprotein complexes on the inner mitochondrial membrane, referred to as nucleoids [Satoh and Kuroiwa 1991; Wiesner et al. 1992].

Maternal inheritance is, with few exceptions, the rule in all animal species [Birky 2001]. There are various ways the paternal inheritance is prevented: the complete lack of mitochondria in crayfish sperm, the mitochondria not entering the egg in the case of the tunicates [Birky 1995], or selective degradation of paternal mitochondria. The latter has been identified in mice [Kaneda et al. 1995], honeybees [Meusel and Moritz 1993], humans [Sutovsky et al. 2004], and, more recently, in *C. elegans* as postfertilization autophagy of sperm organelles [Al Rawi et al. 2011]. Despite this mechanism of selective degradation, the recognition of the paternal mitochondria can fail, leading to leakage. This phenomenon has been documented in mammals, birds, reptiles, fish, mollusks, nematodes, trematodes, and arthropods [e.g., Gyllensten et al. 1991; Jannotti-Passos et al. 2001; Schwartz and Vissing 2002; Kvist 2003; Fontaine et al. 2007; Wolff and Gemmell 2008]. The exception from the maternal inheritance is found in several species of bivalves for which a doubly uniparental inheritance mechanism has been characterized [Zouros et al. 1992, 1994]: female offspring inherit the mtDNA from their mother and the male offspring inherit the mtDNA from both parents, leading to heteroplasmic males.

Heteroplasmy, the coexistence of different mtDNA molecules in the same cell, can be of two types: length heteroplasmy (different nucleotide lengths) or site heteroplasmy (different nucleotide compositions). It was proposed that during mitotic divisions, a bottleneck in the number of mitochondrial genomes accompanies the oogenesis, and only a small subset of mtDNA molecules are amplified and transmitted to the offspring [Jenuth et al. 1996; Marchington et al. 1998]. Later studies showed, however, that a genetic bottleneck occurs during folliculogenesis, as a result of replication of a subpopulation of mtDNAs [Wai et al. 2008]. This way, mutations that escape the filter for purifying selection during the oogenesis are rapidly segregated and exposed to selection at the level of the organism in individual maternal lineages. This fact may be relevant in understanding the evolutionary dynamics of mitochondrial genomes [Koehler et al. 1991; Poulton et al. 1998]. When the number of mitochondria is reduced, the choice in which mitochondria are subsequently amplified is either by random choice of molecules, or on a more systematic bias basis. All these may result in rapid changes in the frequency of different mitochondrial haplotypes over generations. If the heteroplasmy is stable, it may provide useful additional information for defining haplotypes, and resolving further the relationships among individuals at a population level. A famous example of such a case

was during the identification of the remains of Russia's last Tsar, Nicholas II, and his family [Ivanov et al. 1996].

Recombination at the level of the mtDNA has been proven experimentally, by studies of natural populations of several animal species [Lunt and Hyman 1997; Ladoukakis and Zouros 2001; Hoarau et al. 2002; Burzynski et al. 2003; Gibson et al. 2007; Sammler et al. 2011], and detailed biochemical and molecular studies have shown that animal mitochondria possess, indeed, the necessary enzymes for recombination [Thyagarajan et al. 1996]. However, the indirect evidence of recombination based on human mtDNA sequence analysis and linkage disequilibrium [Awadalla et al. 1999; Eyre-Walker et al. 1999a; Hagelberg et al. 1999] have been proven wrong. Shortly after the first publications some discussions on whether recombination actually occurs in human mitochondria were carried in Eyre-Walker et al. [1999b] and Macaulay et al. [1999]. The methodological settings and specific dataset [Piganeau and Eyre-Walker 2004], failure to detect mutations hotspots [Galtier et al. 2006], and alignment errors [Hagelberg et al. 2000] were considered as reasons for false detection of recombination. The confounding effects of recombination on phylogenetic reconstruction was revealed in two complementary studies by Schierup and Hein [2000] and by Posada and Crandall [2002], where they have shown that failure to account for recombination can seriously mislead the population genetic and phylogeographic inferences of a study organism.

There are, however, constrains for the detection of mtDNA recombination in natural populations. The strict maternal inheritance means that recombination would occur in homoplasmic cells, making the detection of recombinants impossible; it is only in heteroplasmic cells that recombination can be detected [Rokas et al. 2003]. Heteroplasmy arises by inheritance of mtDNA from both parents, or by mutations in the mitochondrial genome of germ-line cells. In animals it is rare and recombination leading to new haplotypes is expected to be even more so. Furthermore, the $\mathrm{sperm} : \mathrm{egg}$ mitochondria ratio in the one-cell zygote is very small (perhaps $1 : 10^4$), which makes the occurrence of recombination and the detection of its products highly susceptible to random drift. On the other hand, it is difficult to demonstrate strict maternal inheritance in natural populations because of the large sample sizes required [Birky 2001]. The recombination issue will be raised again in Section 2.4.

2.2.2 Mitochondrial DNA Neutral Evolution

Effective neutrality is a major assumption underlying the analyses of mtDNA population data that has received a lot of criticism. Mitochondria are the site of the terminal stage of the cellular respiration, where most of the ATP is produced, and malfunction of the fundamental complexes would be lethal or, at least, could severely affect the fitness of the organism. Therefore, mtDNA is generally thought to evolve primarily under constant purifying selection. However, the directional selection has gained wide acceptance [Gerber et al. 2001; Bazin et al. 2006]. It is assumed that mtDNA undergoes neutral or deleterious mutations, which are rapidly removed by purifying selection without much affecting the diversity at linked sites. The observable variation would therefore reflect neutral processes only, in agreement with the neutralist theory of molecular evolution [Kimura 1983]. Most studies confirm the neutral evolution of the mtDNA. However direct experiments and statistical tests on model and non-model species, like *Drosophila* [Rand 2001], snakes [Castoe et al. 2008], agamid lizards [Castoe et al. 2009], sharks [Martin et al. 1992], primates [Grossman et al. 2004] and nematodes [Blouin 2000] have detected positive selection during mitochondrial protein-coding gene evolution.

Bazin et al. [2006] conducted a meta-analysis study on more than $1,600$ animal species, showing that the average within-species level of mtDNA diversity is similar across animal phyla, in contrast to nuclear loci that show higher average diversity in, for example, invertebrates than in vertebrates. This conclusion is consistent with the theoretical relationship between population size and genetic diversity. The authors argued that the mitochondrial-specific lack of population size effect can be explained by recurrent selective sweeps (adaptive evolution) in large populations, in agreement with the "genetic draft" model proposed by Gillespie [2000]. The largely nonrecombining nature of the mtDNA makes it highly sensitiv to genetic hitchhiking. According to this, recurrent sweeps affect mtDNA in species with large population sizes, leading to frequent drops in diversity at the whole genome level [Galtier et al. 2009]. Several selective mechanisms can account for adaptive-like patterns:

- ⋆ *selection at the host level*, in which case some mitochondrial variants might be more advantageous: e.g., by inducing more efficient or flexible energetic metabolism [Dowling et al. 2008]
- ⋆ *"selfish" mitochondrial mutations*, that favour the transmission of the mtDNA to the next generation irrespective of the fitness of the host, e.g., the *petite* mutant in *Saccharomyces cerevisiae* [MacAlpine et al. 2001]

⋆ *hitchhiking*, when mitochondria occur within the cell alongside other genetic elements that are also inherited maternally, e.g., maternally inherited symbiotic bacteria *Wolbachia* [Jiggins 2003, and others]

During the course of their evolution, mitochondrial genomes exhibited an extensive reduction in size, indicating a long history of selection. This reduction in the number of the coding genes include: (a) the loss of the genes that became obsolete in the setting of a specialized organelle, such as those involved in nucleotide, lipid, and amino acid biosynthesis [Gray 1992; Berg and Kurland 2000], (b) the elimination of the genes with redundant presence in the nuclear genome—the products encoded by the nuclear genome are used for mitochondrial functions [Gray 1992], like for mitochondrial transfer RNAs (tRNAs) [Duchêne et al. 2009], and (c) the migration of certain genes towards the nucleus, their products being afterwards imported in the mitochondria, such as the genes coding for the complexes involved in the oxidative phosphorylation.

Selection has, no doubt, shaped the evolution on mitochondrial genome, this miniscule piece of genetic information that has played an important role in the study of evolution, ecology, population genetics, and population dispersal.

2.2.3 Clock-like Evolution of the Mitochondrial Genome

It has been shown that the departure from homogeneous rates can be very strong. Molecular phylogenies derived from the analysis of complete or partial mtDNA data have remodeled the classical view of organismal relationships based mainly on morphological data. With this remodeling, inconsistencies started to surface between the morphological and molecular data, and, also, between nuclear DNA (nDNA) and mtDNA molecular data. Mitochondrial DNA has been extensively used as phylogenetic marker for recent time scales (within the past 5–10 million years) and this is due to its relatively high substitution rate comparing to nDNA [Brown et al. 1979]. The famous 2% per million years divergence was often considered as a reasonable reference in the absence of relevant fossil data in mammals and more generally in vertebrates [Moritz et al. 1987]. There are several reasons thought to be responsible for a fast evolution rate for the mtDNA: (a) it is less protected by proteins than the nDNA, (b) it is associated with the inner membrane of the mitochondria, where reactive oxygen species are generated, and (c) it appears to have a less-efficient repair mechanism than the nDNA.

Studies by Gissi et al. [2000], Kumar and Subramanian [2002], and Bininda-Emonds

[2007] showed that in vertebrates the substitution rate has a two- to threefold difference between fast-evolving and slow-evolving lineages, supporting the notion that the mtDNA evolutionary rate is little variable across taxa. According to Galtier et al. [2009] the substitution rate estimation is affected by multiple substitutions and those would not be visible because the aforementioned studies used a small number of very diverged sequences. Studies based on recent diverged species, large fossil record, and more than $1,500$ *cytb* gene sequences have shown that the substitution rate varies 30-fold between fast and slow-evolving bird lineages [Nabholz et al. 2009], and 100-fold between mammalian lineages [Nabholz et al. 2008]. Three main hypotheses have been proposed to explain the variation of the substitution rate in mtDNA, the generation time, the metabolic rate, and the longevity hypotheses, but there were arguments both supporting them [Nabholz et al. 2008, 2009] and rejecting them [Kumar and Subramanian 2002, and others].

2.3 Mitochondrial Genomes

«Publication of the complete sequence of the $16,569$-base pair human mitochondrial DNA on p.457 of this issue of *Nature* is a landmark in biology. It comes only 15 years after mtDNA from vertebrates was found to consist of rather small DNA circles and only 6 years after Sanger and his co-workers developed the sequencing methodology to tackle such large molecules »

Piet Borst and Leslie A. Grivell, *Small is beautiful—portrait of a mitochondrial genome*, Nature, 1981

Most organelles are thought to harbour single circular genomes similar to those of prokaryotes, but many organelle genomes exist as linear head-to-tail concatemers produced by rolling-circle replication [Bendich 2004]. Linear mitochondrial genomes have been documented in some fungi [Forget et al. 2002], parasitic arthropods [Shao and Barker 2007], and some plants [Oldenburg and Bendich 2004]. Furthermore, there seem to be deviations from the single organelle molecule. The kinetoplastid protists (e.g., *Trypanosoma*) mitochondrion consists of thousands of circular molecules that need post-transcriptional editing [Morris et al. 2001]. The mitochondrion of the ichthyosporean *Amoebidium parasiticum* consists of hundreds of short linear fragments with an approximative total size of 300 kb [Burger et al. 2003]. The potato cyst nematode *Globodera pallida* has four mitochondrial subgenomes [Gibson et al. 2007], and the phytonematode *Meloidogyne javanica* mtDNA is characterized by subgenomic circles thought to be produced by the breakage and rejoining of the progenitor mtDNA [Lunt and Hyman 1997].

According to the Organelle Genome Resources Database from NCBI as of October 2011, the size of the mtDNA in most eukaryotic phyla ranges from 15 to 20 kb, but

there are some notable exceptions. The smallest mtDNA molecule sequenced belongs to *Plasmodium falciparum*, the malaria parasite, with just under 6 kb in length and only three protein coding regions [Conway et al. 2000]. The largest eukaryotic mtDNA belongs to *Arabidopsis thaliana*, the thale cress, with over 360 kb and 57 genes, that cover only about 10% of the genome [Unseld et al. 1997]. More recent studies have identified that the largest mitochondrial genome belongs to *Cucumis melo*, the melon, with more than 2.74 Mb, but the assembly is not finalized [Rodríguez-Moreno et al. 2011]. In general, plants tend to have much larger mtDNAs than animals, although they do not have proportionately greater coding capacity [Gray et al. 1999; Lang et al. 1999]. Around 90% of the mtDNA of plants is composed of intergenic and various repeated regions [Kitazaki and Kubo 2010]. In the Viridiplantae clade (green algae and land plants), only three mitochondrial genomes are under 20 kb, of which *Polytomella capuana* (13 kb), the colorless green alga [Smith and Lee 2007], is the smallest one. The differences in size are mostly due to marked variations in the length and organization of the intergenic regions.

As mentioned before, many of the ancestral mtDNA genes have been moved to the nDNA. As the organelles exist in many copies per cell, the advantage of having the genes moved into the nDNA would be that, during the cell division in diploid organisms, each gene has to be copied only twice, rather than hundreds of times. Most of the genes used by the mitochondria are not located within its genome, but only in the nucleus, so the nucleus controls the multiplication rates of the organelles. With this control over the organelles, a potential intracellular conflict is avoided. Nevertheless, the mtDNA still retained some coding potential. The energy-producing function of organelles causes production of oxygen radicals, highly reactive and mutagenic molecules. As the organelles themselves are in control of the expression of the genes encoding the components responsible for their electron transport chain, the feedback is rapid and precise [Allen 2003]. Another possible reason for keeping some of the coding capacity is that, once any change in the genetic code has taken place, further transfer of genes to the nucleus would lead to the synthesis of erroneous proteins. As some protein-coding genes had remained in the organelles, it was also necessary for them to retain the genes coding for tRNA and rRNA. If this explanation is correct, it is partly a frozen accident that particular protein-coding genes were retained: they are those genes that happened to be still in the mitochondria when the code changed [Maynard Smith and Szathmáry 1995]. The mere fact that an import machinery actually exists in the mitochondria supports a hierarchical loss of the genes from the mitochondria to the nucleus. As the mtDNA encodes only certain genes responsible for the electron transport chain, then some genes might be

easier transferable than others. Features that rendered some genes unsuitable to be transfered include the hydrophobicity of the proteins (e.g., COX1) or their large size (e.g., small-subunit and large-subunit of the rRNAs). Both nuclear and mitochondrial genes contribute to mitochondrial protein function and both genomes interact to affect the synthesis and assembly of mitochondrial proteins [Poyton and McEwen 1996].

2.3.1 Animal Mitochondrial Genomes

The Organelle Genome Resources Database contains $2,663$ records of completely sequenced mitochondrial genomes, the majority of them ($2,436$ records) belonging to Metazoa. With few exception, animal mtDNAs are 15–20 kb circular molecules (as seen in Figure 2.1). The smallest complete mtDNA of about 8 kb in length, five protein coding genes and 14 structural RNAs, belongs to *Anaticola crassicornis*, the slender duck louse [Cameron et al. 2011]. The largest mtDNA belongs to the placozoan *Trichoplax adhaerens*, with 40 kb in length, 20 protein coding genes, 27 structural RNAs, intragenic spacers, introns, and unidentified ORFs [Dellaporta et al. 2006]. In spite of being more conserved than the plant mtDNAs, there is significant variation in size and gene content between and also within the major metazoan groups (see standard deviation bars in Figure 2.1).

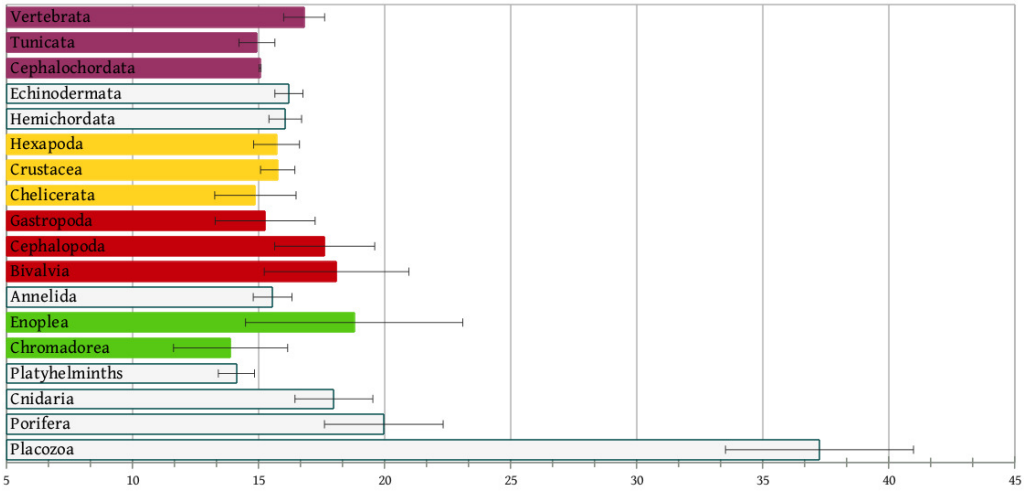


Figure 2.1 – Mean length of complete mtDNAs. The mean length (kb $\pm$ standard deviation bars) of complete mtDNAs calculated for the major metazoan groups ($2,350$ from the $2,436$ metazoan entries have been sampled for this diagram). Bordeaux–Chordata, Yellow–Arthropoda, Red–Mollusca, Green–Nematoda. Figure after Gissi et al. [2008], updated according to the Organelle Genome Resources Database as of October 2011.

Codon Usage

The analysis of amino acid sequences of mtDNA genes from different organisms carried out in the early 80's proved that mtDNA uses a different genetic code than the universal one [Anderson et al. 1981]. Today, ten different mitochondrial codes have been identified, among which a small number of codons change the encoded amino acids (Figure 2.2).

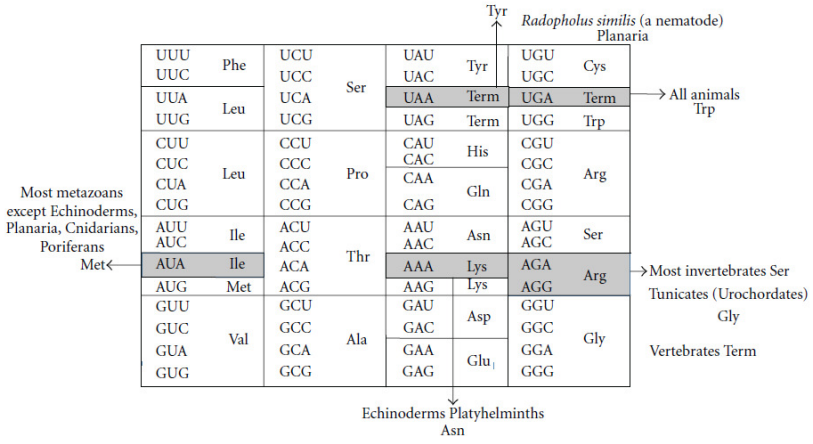


Figure 2.2 – Genetic codes. Universal genetic code (inside the box) and variations in animal mitochondrial genetic code (outside). Term: termination codon. Figure from Watanabe and Yokobori [2011].

Important changes are the ones involving the termination codons (in all metazoans, mtDNA UGA specifies tryptophan rather than a STOP codon), and alternative initiation codons. In vertebrate mtDNA, the codons AUU, AUC, AUA and GUG are used for initiation and in invertebrate mtDNA, the codon UUG, in addition to the ones used in the vertebrate mtDNA, is also used as initiation codon. Codon reassignments can be caused by changes in the tRNAs. During this process, the organisms may be under a great selective disadvantage. The reassignment introduces an amino acid change in every protein where this codon occurs. The situation is worsened when STOP codons are involved, yet they constitute a majority of the observed reassignments. A more peculiar way of initiation is found in *Drosophila yakuba* and *D. melanogaster*: the codon AUA is out of frame with the rest of the *cytb* gene, so the initiation appears to require the reading of a AUAA quadruplet, which would need a specific ribosomal frameshift [de Bruijn 1983], or it would be edited in order to be brought into frame [Fox 1987]. The termination codons seem also to put some additional problems as some protein genes end either in T or TA, rather than a complete translation termination codon like TAA or TAG. These incomplete stop codons are followed immediately by tRNA genes. It has been shown that these codons are completed by post-transcriptional polyadenylation [Ojala et al. 1981].

Gene content

The most frequent mtDNA gene content is 37 genes encoding 13 respiratory chain proteins (ATP synthase: *atp6* and *atp8*, apocytochrome b: *cytb*, cytochrome oxidase: *cox1*, *cox2*, and *cox3*, and reduced nicotinamide adenine dinucleotide ubiquinone oxireductase: *nad1–6*, and *nad4L*), and the structural RNAs—22 tRNAs and two rRNAs (*rrnS* and *rrnL*). However, the gene number varies from 13 genes in the sea walnut *Mnemiopsis leidyi* and the arrow worm *Spadella cephaloptera*, to 53 genes in the western predatory mite *Metaseiulus occidentalis* [Jeyaprakash and Hoy 2007]. The mtDNA of most metazoans is characterized by an asymmetric distribution of the genes between the two strands, allowing the identification of a major and a minor strand, also called heavy and light chain, depending on the number of encoded genes. The distribution of genes between the two strands is directly related to the mechanism of mtDNA transcription [Gissi et al. 2000].

These highly derived diminutive genomes consist of 90%–95% coding DNA exhibiting overlaps between the initiation and the termination codons of some adjacent genes [Lynch 2006b]. The human mtDNA molecule was the first mitochondrial genome ever to be completely sequenced, analyzed, and interpreted with regard to gene content. With few exceptions, mtDNA molecules from all metazoans have proven to exhibit the same gene content, but there is substantial great deal of variation regarding the relative arrangement of these genes [Wolstenholme 1992]. They all lack introns, with the exception of two cnidarians, the coral *Acropora tenuis* [van Oppen et al. 2002] and the sea anemone *Metridium senile* [Beagley et al. 1998], that contain one and two introns, respectively, and the placozoan *Trichoplax adhaerens* [Dellaporta et al. 2006], which contains four introns. These exceptions can reveal potential mechanisms of maintenance of organelle introns.

Lose or acquisition of protein-coding genes is rather infrequent and must be treated with care due to uncertainty in gene annotation or the availability of incomplete mtDNA sequences. This has been the case for Antarctic icefishes of the suborder Notothenioidei, where *nad6* seemed to have been lost [Papetti et al. 2007], but later studies showed that the "missing" gene has been translocated in the control region of the mtDNA [Zhuang and Cheng 2010]. Despite this case, there are indeed several genes that appear to have been lost more frequently than others: *atp6* is absent in all Chaetognatha (arrow worms), and *atp8* has been modified or lost several times independently. It is not known yet, if these losses are related to the transfer of the gene to the nuclear genome, or they are

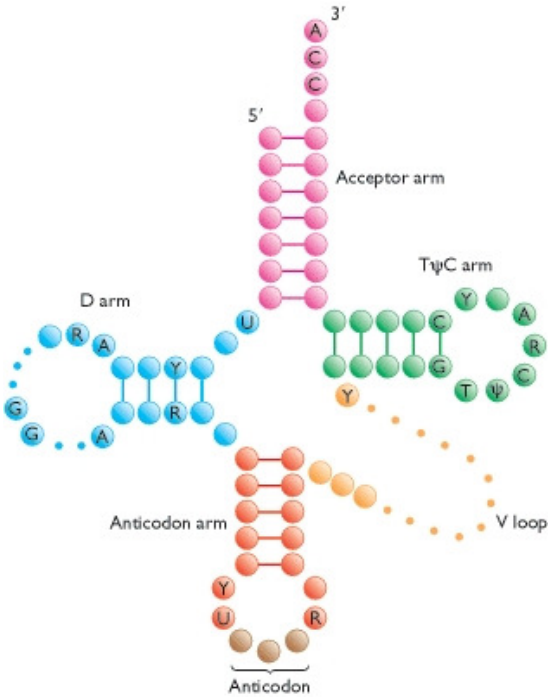


Figure 2.3 – The cloverleaf secondary structure model of a tRNA gene. Letters indicated identify nucleotide or nucleotide combinations that are considered invariant (A, C, G, T, U, ψ (pseudouridine)) or semi-invariant (R (purine) and Y (pyrimidine)). The 3′ terminal CCA nucleotides either are DNA-encoded (most prokaryotic tRNAs) or are added post-transcriptionally (most eukaryotic tRNAs). The smaller dots represent optional nucleotides present in some tRNAs. In standard numbering system based on Sprinzl et al. [1989], the nucleotides of the anticodon are at the positions 34, 35 and 36.

actually due to differences in the ATP synthase structure between taxa [Gissi et al. 2008].

It is considered that the mtDNA encodes for a full set of tRNAs able to decode the modified genetic code mentioned previously. This full set can vary from one taxon to another, but it is nevertheless smaller than the nuclear one. The reduction is achieved by the use of a single tRNA with the nucleotide U in the first anticodon (wobble) position to recognise all codons of a four-codon family [Taanman 1999]. The number of tRNAs encoded in a given mitochondrial genome depends on many factors, like the genetic code, evolutionary history of the taxa, the ability to acquire or lose genetic material, and the capacity to import and use nuclear encoded tRNA molecules. Thus, depending on the mitochondrial genetic code used, the expected set of mitochondrial tRNAs ranges from 24 genes (UGA encodes for Trp rather than STOP) to 22–23 genes (modified meaning of UGA, AUA and AGR codons). The use of a particular genetic code does not lead to the fact that the same number of tRNAs are used to decode the genome. The number of tRNAs used is rather connected with the frequency of losses and duplications of tRNA genes. For example, in the case of the cnidarian *M. senile*, the mtDNA encodes for only two tRNA genes, the rest are presumably imported from the nDNA [Beagley et al. 1998]. Severe deviations from the usual set of tRNA genes have been also identified in other organisms [Beagley et al. 1998], but this is debatable as misannotations have been found in the databases [Gissi et al. 2008].

The primary and secondary structures of tRNAs encoded in prokaryotic DNAs, nuclear

DNAs of protists, animals and plants, and chloroplast DNAs have been highly conserved. In contrast, among tRNAs encoded by metazoan mtDNAs, there is a remarkable diversity of structure (Figure 2.3). The cloverleaf secondary structure is adopted by most vertebrate and invertebrate mitochondrial tRNAs, but in platyhelminths, insects, echinoderms, and some vertebrates, there is variation in the size and sequence of the loops of the dihydrouridine (DHU) and $\mathrm{T}\psi\mathrm{C}$ arms. The extent that the invariant and semi-invariant nucleotides are essential for the folding of standard tRNAs into a functional tertiary configuration varies between the different taxa.

All metazoan mtDNA molecules examined so far contain two rRNA genes (*rrnS* and *rrnL*) that have been identified from sequence comparisons as the homologs of rRNA genes of *Escherichia coli*. Though their duplication is rare, it has, nevertheless, been identified in the bivalve *Crassostrea gigas* [Milbury and Gaffney 2005], in distinct haplotypes of the nematode *Thaumamermis cosgrovei*, in the chigger mite *Leptotrombidium pallidum* [Shao et al. 2005], and in the mosquito parasite nematode *Strelkovimermis spiculatus* [S Tang and B Hyman, direct submission to GenBank].

Intergenic Regions

Although compact in its organization, the animal mtDNA contains some intergenic, noncoding regions. The most extensive one is a region typically of around 1 kb in length and bound by tRNA genes, called D-loop, displacement-loop, or control region, that has been shown to contain sequences essential for the initiation of transcription and replication [Wolstenholme and Jeon 1992; Clayton 2003]. It is in general poorly conserved between species and varies greatly in size and nucleotide composition.

2.3.2 Nematode Mitochondrial Genomes

The Organelle Genome Resources Database contains 58 complete mitochondrial genomes from Pseudocoelomata, 54 of which belong to the phylum Nematoda, one to the phylum Acanthocephala (*Leptorhynchoides thecatus*, the thorny-headed worm) and three to the phylum Rotifera (*Rotaria rotatoria* and *Brachionus plicatilis*). Table 2.1 lists several molecular size estimates for the mtDNA of representative nematodes. From the entire set of completely sequenced nematode mtDNAs, two enoplean nematodes have the extreme sizes: *Xiphinema americanum*, the American dagger nematode, is the smallest with $12,626$ bp, while *Romanomermis culicivorax* is the largest with $26,194$ bp. Among

the chromadoreans, *Heterorhabditis bacteriophora* has the largest sequenced mtDNA, and *Angiostrongylus cantonensis*, the rat lungworm, the smallest.

Genetic code

According to the Invertebrate Mitochondrial Code, UGA specifies tryptophan rather than STOP codon, AUA specifies methionine and not isoleucine, AGA and AGG specify serine instead of arginine and AUG, AUU, AUA, or UUG codons can serve as initiation codons [Okimoto et al. 1992, 1990] (see also Figure 2.2). Across Pseudocoelomata, there is an obvious sampling bias, Phylum Nematoda being far overrepresented. With respect to the genetic code used in translation of the mtDNA, there is one exception from the use of the Invertebrate Mitochondrial Code and that one belongs to *Radopholus similis*, the burrowing nematode: the codon UAA has been found to code for tyrosine rather than STOP [Jacob et al. 2009] (therefore, *R. similis* is using the Alternative Flatworm Mitochondrial Code).

Gene content

The phylum Nematoda is a little apart in the metazoan group: there are only 12 protein-coding genes, *atp8* being absent from all nematode mtDNAs, except from *Trichinella spiralis*, the trichinellosis nematode [Lavrov and Brown 2001]. Other pecularities with respect to gene content are in the cutworm parasitic nematode *Hexamermis agrotis* that has the gene *atp6* triplicated and in the mosquito parasites *R. culicivorax* and *R. iyengari* that have a *nad3* duplication and an *atp6* triplication, respectively. In the case of the nematode *T. cosgrovei*, both the genes *nad6* and *nad4* are duplicated as nonidentical, putatively nonfunctional partial copies [Tang and Hyman 2007]. With the exception of the nematodes from Clade I *sensu* Blaxter et al. [1998], the nematode mtDNA is transcribed unidirectionally.

Incomplete termination of protein-coding genes is very commonly found in nematode mtDNAs, including *T. cosgrovei* [Tang and Hyman 2007], *Anisakis simplex* [Kim et al. 2006], *Toxocara* species [Li et al. 2008], *Cooperia oncophora* [Van Der Veer and De Vries 2004], *C. elegans* [Okimoto et al. 1992], *Pristionchus pacificus* [Molnar et al. 2011], and many others. If incomplete terminations are not accepted, and instead the complete termination codon (UAA or UAG) is preferred, then some gene overlaps would be required. This is unlikely, though, because the resulting gene size would far exceed the

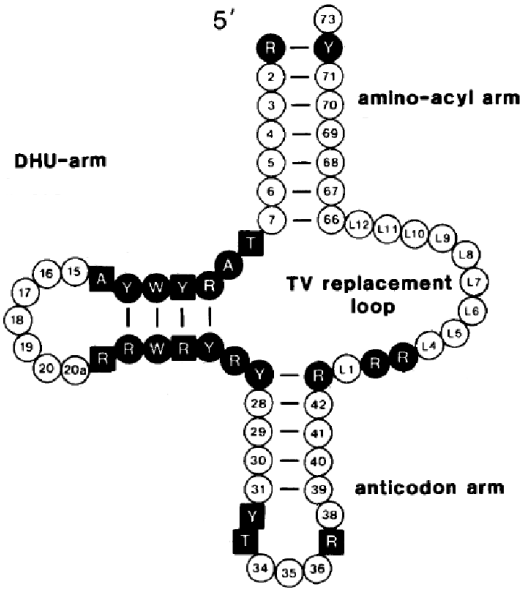


Figure 2.4 – The consensus mitochondrial tRNA secondary structure in nematodes. The number of nucleotides shown in the DHU and the TV-replacement loops are the maximum numbers observed. Letters in black squares and circles identify nucleotides invariant or semivariant in nematode mitochondrial tRNAs. The numbering system used (1–43 and 66–73, given in white circles) reflects the conventional numbering system used for yeast tRNA-Phe [Sprinzl et al. 1989]. L1–L12 denote the maximum of 12 nucleotides that occur in the TV-replacement loop. This figure is modified from Wolstenholme [1992].

lenght characteristic of mtDNA genes in nematodes.

In nematodes, 22 tRNAs are sufficient to decode the mitochondrial genes. There are, however, several species that have more tRNAs due to duplicated segments. Such situations can be seen in *R. culicivorax*, *S. spiculatus*, *T. cosgrovei*, the Guinea worm *Dracunculus medinensis*, and several others. One single nematode species, *X. americanum*, shows an apparent constriction of the tRNA complex to 21 components. The secondary structure of the tRNAs in nematodes is a "bizarre" one (Figure 2.4), as the genes are smaller than their metazoan homologs and therefore the standard arms, as seen in Figure 2.3, are shortened or replaced [Wolstenholme et al. 1987].

In general there are two rRNA genes, *rrnS* and *rrnL*, with some exceptions that have been mentioned across this chapter and in Table 2.1.

Table 2.1 – Mitochondrial Genome Size and Gene Content Overview for Representative Nematodes [modified after Molnar et al. 2011]

	Clades				Gene Content			
Nematode	Blaxter et al. [1998]	van Megen et al. [2009]	Size (bp)	Nonconding Nucleotides (bp)[a]	PCGs[*]	tRNAs	rRNAs	Reference
Pristionchus pacificus	V	9A	15,955	2,772	12	22+tRNA^{Sup}	2	Molnar et al. [2011]
Caenorhabditis elegans	V	9A	13,794	575	12	22	2	Okimoto et al. [1992]
Ancyclostoma duodenale	V	9B	13,721	513	12	22	2	Hu et al. [2002]
Necator americanus	V	9B	13,604	387	12	22	2	Hu et al. [2002]
Steinernema carpocapsae	IV	10A	13,925	696	12	22	2	Montiel et al. [2006]
Strongyloides stercoralis	IV	10B	13,758	575	12	22	2	Hu et al. [2003]
Meloidogyne javanica	IV	12B	20,565	7,000	12	22	2	Okimoto et al. [1991]
Radopholus similis[b]	IV	12B	16,791	3,904	12	22	2	Jacob et al. [2009]
Enterobius vermicularis	III	8A	14,010	675	12	22	2	Kang et al. [2009]
Onchocerca volvulus	III	8B	13,747	357	12	22	2	Keddie et al. [1998]
Ascaris suum	III	8B	14,284	1,048	12	22	2	Okimoto et al. [1992]
Thaumamermis cosgrovei[c]	I	2A	20,012	749[d]	12[d]	22[d]	2[d]	Tang and Hyman [2007]
Trichinella spiralis[c]	I	2A	21,000–24,000	8,098–10,098	13	22+tRNA^{Met2}	2	Lavrov and Brown [2001]
Xiphinema americanum[c]	I	2C	12,626	305	12	17	2	He et al. [2005]

[*] Note–PCG, protein-coding genes

[a] Sum of the bp included in the D-loop and other noncoding nucleotide regions longer than 10 bp

[b] UAA-tRNA^{STOP} to UAA-tRNA^{Tyr} reassignment

[c] The mtDNA is transcribed bidirectionally

[d] Excluding the hypervariable region that contains putative *rrnS* and tRNAs genes

2.4 Accumulation of Mutations

Mutation accumulation surveys in plant, animal and fungal mitochondrial and nuclear genomes consistently found that mitochondrial genomes [Lynch 1997; Lynch and Blanchard 1998] and nuclear genomes of asexuals [Normark 2000] and selfers [Weinreich and Rand 2000; Bustamante et al. 2002; Glémin et al. 2006] accumulate deleterious mutations at higher rate than nuclear genomes of outcrossing sexuals [Charlesworth and Wright 2001]. There are a least three factors that promote elevated mutation rates in animal mitochondria, comparing to their nuclear genomes [Lynch 2006a]:

1. Mitochondria are the sites of aerobic respiration, generating high levels of reactive oxygen species and, consequently, an internal environment with exceptional potential for causing DNA damage: deamination of cytosine (C) to uracil (U), causing $\mathrm{C} : \mathrm{G} \rightarrow \mathrm{T} : \mathrm{A}$ transitions and the oxidative modification of guanine (G) to 8-oxo-guanine, causing $\mathrm{C} : \mathrm{G} \rightarrow \mathrm{A} : \mathrm{T}$ transversions.
2. MtDNA is frequently replicated within non-dividing cells (in contrast to nDNA), increasing the opportunities for replication errors per cell cycle: the base-misincorporation rate (before proofreading) is 10^3–10^4 times higher than in the nuclear genomes.
3. Mitochondrial genomes do not encode for DNA repair proteins – there is evidence that the nucleotide-excision repair mechanism has been lost, at least from mammalian lineages.

The influence of the mitochondrial bottleneck process was higly debated. On one hand, the bottleneck increases the variance within organisms or within populations, thus opposing to mutation accumulation by increasing the efficiency of selection against deleterious mutations [Rand 2008; White et al. 2008]. On the other hand, the genetic drift due to bottlenecking has been proposed to actually explain the transmission of deleterious mtDNA mutations in humans [Chinnery et al. 2000].

2.4.1 Direct Estimation of the Mutation Rates

The Mutation Accumulation Lines Experiment

Our knowledge about the nature of new genetic variation originating from spontaneous mutations has come from mutation accumulation (MA) experiments. In a typical MA experiment (Figure 2.5), mutations are allowed to accumulate over the whole genome for many generations, in replicated inbred lines. Typically, the animals are reared in constant

environment with plenty of food, and selection is expected to be dramatically reduced so that the majority of mutations (all but the most deleterious ones) become fixed or are removed at random.

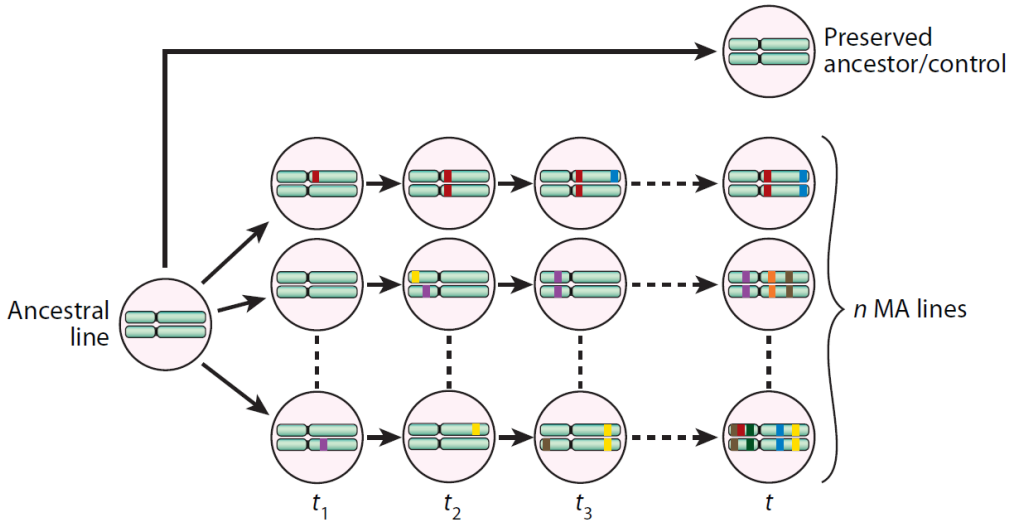


Figure 2.5 – A typical mutation accumulation experiment in a selfing diploid organism. An acestral line is split into n MA lines, which are allowed to accumulate mutations for t generations. Mutations are heterozygous when they first appear and becomse fixed or lost in subsequent generations. After t generations, lines have accumulated different sets of mutations. Mutations are represented as colored blocks within chromosomes. Figure from [Halligan and Keightley 2009].

Mutation accumulation experiments on *Drosophyila* have been pioneered by Muller in 1920's and continued by Mukai in the 60's. Since then, many MA studies have been carried out in a variety of organisms (bacteria, yeast, *Arabidopsis*, *Daphnia*, *Caenorhabditis*), and lines have been assayed for diverse traits. In this section I will summarize only those studies that had as focus the direct estimation of mutation rates in mitochondrial genomes (Table 2.2).

Table 2.2 – Mitochondrial genome mutation rates in MA lines experiments

Species	μ_{bs} [a]	Study
Caenorhabditis elegans	9.7×10^{-8}	Denver et al. [2000]
Saccharomyces cerevisiae	1.2×10^{-8}	Lynch et al. [2008]
Drosophila melanogster	6.2×10^{-8}	Haag-Liautard et al. [2008]
Caenorhabditis briggsae	7.2×10^{-8}	Howe et al. [2010]
Daphnia pulex – sexual lines	2.0×10^{-8}	Xu et al. [2011]
Daphnia pulex – asexual lines	4.3×10^{-8}	Xu et al. [2011]

[a] base substitution mutation rates, calculated per site per generation

Caenorhabditis elegans

The details of the initiation of the mutation accumulation lines in *C. elegans* were published in Vassilieva and Lynch [1999]. The mutation-accumulation experiment is initiated with a single individual derived from a wild-type strain. The strain has been previously maintained by inbreeding for dozens of generations before the experiment was conducted. A single lineage was further inbred by self-fertilization for several generations before initiating the study with the selfed offspring from a single hermaphrodite. Thus, the amount of residual heterozygosity in the base individual should have been no more than expected under selection-mutation balance in a line containing a single individual [Lynch and Milligan 1994]. From the F3 descendants of the base individual, 100 mutation-accumulation lines were initiated, while many thousands of additional animals were cryopreserved and stored at $-80\,^{\circ}\mathrm{C}$. The frozen stock of the original genotype serves as a control for all subsequent assays of the experimental lines.

The nematodes are cultured and handled on NGM agar at $20\,^{\circ}\mathrm{C}$ and maintained with *E. coli* OP50 as food source. To minimize the efficiency of natural selection against new mutations, each of the lines are propagated across generations as a single random worm from the L3 to young adult stages, picked from the middle of the parental reproductive period. Accidental losses of the experimental lines are prevented by maintaining two previous generations at $15\,^{\circ}\mathrm{C}$, as potential backups.

Denver et al. [2000] estimated the mutation rate and pattern in the mtDNA of 74 *C. elegans* MA lines at the 214^{th} generation. They identified 26 mutations which gave the total mutation rate estimate of 1.6×10^{-7} ($\pm 3.1 \times 10^{-8}$) per site per generation. No hotspot and no lines were found predisposed to bear mutations. The mutation rate for base substitution mutations is 9.7×10^{-8} per site per generation. From the 26 mutations, 10 were insertion-deletion (indel) mutations mostly associated with homopolymeric or repetitive regions. Similar mutation rates were observed for *C. briggsae* mtDNA (1.1×10^{-7} or 7.2×10^{-8}, depending on the strain used) [Howe et al. 2010].

Saccharomyces cerevisiae

In yeast, the propagation of the MA lines is done by passing each line through single-cell bottlenecks on a 3- to 4-day cycle. Each cycle is initiated by picking a random colony and streaking onto rich YPD medium. Random propagation is imposed by overlaying each plate on a template with a marked target and selecting the colony closest to the target.

Lynch et al. [2008] estimated the mutation rate for base substitutions at 12.2×10^{-9} per site per cell division and 7.5×10^{-9} per site per cell division as the indel mutation rate. Interestingly, the yeast mtDNA seems to have a much higher mutation rate than its nuclear genome ($37\times$), much higher than vertebrates and invertebrates exhibit ($19\times$ for vertebrates and $8\times$ for invertebrates) [Lynch 2007]. All the base substitutions mutations identified were $\mathrm{A:T} \rightarrow \mathrm{G:C}$, despite the strong $\mathrm{A/T}$ bias of the genome.

Drosophila melanogaster

A weakness of the MA line approach in *D. melanogaster* is the difficulty of maintaing a control in species with no resting stage to the life cycle [Houle and Nuzhdin 2004]. Also, cryopreservation results show that the survival rate of recovered embryos is at only about 5%. An additional trouble while propagating the MA lines for *Drospophila* comes from the mode of reproduction. In order to minimize selection within vials, single pairs are allowed to lay eggs for a short while and offspring that initiate the next generation are chosen at random with respect to development time. The MA lines can be propagated by brother-sister mating (full-sib or half-sib mating) or/and double first-cousin mating.

In a very elegant study, Haag-Liautard et al. [2008] estimated the mutation rate of *D. melanogaster* mtDNA at 6.2×10^{-8} per site per fly generation for base substitutions, and 1.6×10^{-8} per site per generation for indels. Most of the base substitution mutations appeared to be $\mathrm{G:C} \rightarrow \mathrm{A:T}$, in accordance with the strong $\mathrm{A/T}$ biased mtDNA genome of *Drosophila*.

Daphnia pulex

Daphnia pulex reproduces by cyclical parthenogenesis (sexual), but numerous obligate asexual lineages have arisen, which provides a great system to examine mutation rate variation. Although, mtDNA is inherited "asexually" in both sexual and asexual species, the mitochondrial-nuclear crosstalk might influence the mtDNA mutation rate.

The MA lines of *D. pulex* are maintained in benign laboratory environment and are propagated asexually, with a generation time of 10–30 days. Each line is forced through bottleneck each generation by randomly picking a single individual to continue the line, simultaneously with two additional females from the same brood kept as backups. Xu et al. [2011] estimated the base substitution mutation rates at 2.0×10^{-8} and 4.3×10^{-8} per site per generation for the sexual and asexual lines, respectively.

2.5 Instead of Conclusion

We've seen that the mitochondrial genome architecture has many faces, with gene fusions, gene-in-pieces, gene simplification, and huge size differences between the taxa [Burger et al. 2003]; and this is just to mention the "morphological" aspects. When it comes to genetics, there is an entire list to consider. As written at the beginning of the chapter: mtDNA is well studied and well conserved, but there is so much variation...

2.5.1 Reason for This Chapter

or Why This Is Directly Relevant to My Line of Research

In this chapter, I made a relatively extensive introduction concerning the mitochondria and their mitochondrial DNA, including the types of mutations and some of the reasons these accumulate in the mtDNA. In section 2.1, I presented a very short overview on the initial theories about the origins of the mitochondria as organelle. Section 2.2 gave some insights into how the mtDNA is inherited, how and why heteroplasmies exist, whether or not recombination is affecting studies of population genetics, or the neutral evolution of the mtDNA is really neutral and some reasons to worry about it. In this section the concept of clock-like evolution is introduced alongside the reasons that can lead to consider a non-clock-like evolution. MtDNA is used to calculate divergence time divergence populations, strains, or species and in this section I refer the problems that can come with estimates from this locus. In section 2.3, I presented an overview of the size and types of mtDNAs, with emphasis on animal mtDNA in general and nematode mtDNA in particular, and the reasons for which the mtDNA has only few genes. The chapter ended with section 2.4 that helps us understand how and why mutations accumulate in the mtDNA and how mutation-accumulation line experiments contribute to directly estimate the mutation rates.

This is in accordance with the research article presented in this thesis which includes the description of the mtDNA of the nematode *Pristionchus pacificus*, mutation rates estimated via the mutation accumulation lines experiment, the spectrum of mutations found, a comparative view on the mtDNA composition between several strains of *P. pacificus*, and, finally, an insight into the minimal divergence time estimates between the strains. All this work will be described in detail in Chapter 4.

Chapter 3

Microsatellites

simple sequences with complex evolution – Hans Ellegren

This chapter will describe microsatellites, very dynamic stretches of repetitive DNA sequences found in the genomes of all organisms and organelles. The description will not be exhaustive over the knowledge of these genomic elements, but it will present the necessary properties for understanding their apparition and their mutation dynamic. The first sections will focus on the definition and the genomic context in which these elements are found in. I will discuss the types of microsatellites and their distribution in the genomes, the types of mutations that occur at these loci, and their mutations rates. This chapter is concluded with a paper submitted for publication in 2012.

Few genetic markers, if any, have found such widespread use as the microsatellites [Ellegren 2004]. They are used as markers in population genetics from the 1980's, especially for mapping chromosome segments, for the identification of individuals, for tracking the biological history of populations, and for paternity and maternity during sexual selection, due to their great advantage of neutrality [Chambers and MacAvoy 2000]. Short-repeated DNA sequences of all motif lengths have been shown to occur in eukaryotic genomes far more frequently than it would be expected by chance [Hancock 1999] and their dynamic has been well studied. But, why are microsatellites so abundant, and why are they so polymorphic? What mechanism do they mutate by? What governs the balance between genesis and expansion of simple repetitive arrays? Why do microsatellites rarely reach really long lengths? How do they evolve? [Ellegren 2004]. I will answer these questions in this chapter.

3.1 Microsatellites Genome Biology

«There is a good body of evidence showing that microsatellite evolution is not as simple as the synonym for microsatellites – simple sequences – might suggest.»

Christian Schlötterer (Evolutionary geneticist) *Genome evolution: Are microsatellites really simple sequences?* Current Biology, 1998.

3.1.1 Definition

The microsatellites belong to the group of genomic elements called tandem repeats. They are composed of different numbers of adjacent repeats of a given motif, or repeat unit. They are variable in the number of repeats, which is a consequence of particular mutation processes, the polymerase slippage and unequal recombination [Hancock 1999]. Based on the number of tandem repeats, there are three classes of satellite DNA that can be distinguished:

⋆ *Satellites* — tandem repeats of variable lengths organized in large clusters that can usually reach several megabases, found in the heterochromatic regions of chromosomes, near centromeres and telomeres or on the Y chromosomes [Lohe et al. 1993] and quite stable in array size with populations [Charlesworth et al. 1994].

⋆ *Minisatellites* — tandem repeats of 10–100 bp that can form tracts of more than 500 bp [Abouelhoda et al. 2010], found in euchromatic regions of the genome [Armour and Jeffreys 1992].

⋆ *Microsatellites* are the equivalent to minisatellites, but the tandem repeat units are 1 or 2, to 6 nucleotides (nt) long and form tracts of around 10^2 bp, found widespread all through prokaryotic and eukaryotic genomes [Toth et al. 2000].

For being rather simple things, microsatellites have proved to be rather tricky to describe and classify. The term microsatellite has been historically used only for repeats of the dinucleotide $\mathrm{CA} : \mathrm{GT}$ [Litt and Luty 1989] and a variety of other names have been used to describe tandemly repeated sequences [Hancock 1999]. Such names, like 'short tandem repeats' [STRs; Edwards et al. 1991], 'simple sequence repeats' [SSRs; Haynes et al. 1989], or 'simple repetitive DNA' [SR-DNA; Levinson and Gutman 1987] or 'simple sequence length polymorphisms' [SSLPs; Tautz 1989] are still widely used in the literature. Nakamura et al. [1987] proposed that both minisatellites and microsatellites should be called 'variable number of tandem repeats' (VNTRs). The difference between minisatellites and microsatellites is arbitrary, but it is important to take into consideration

the apparent size difference between them, as it is believed they evolve by different mechanisms [Armour and Jeffreys 1992]. So many names, make the literature research difficult to perform.

3.1.2 Molecular Structure

The common characteristic to all microsatellite definitions is that they have a sequence of nucleotides repeated tandemly. But the definitions of repeat unit size, number of repeats, and complexity of the sequence are still under debate [Ellegren 2004].

Repeat Unit

Microsatellite loci are represented by the general formula $-(\mathrm{X})_n-$ where X is the repeat unit, fixed in the range 1 or 2 to 6 nt and n is the number of repeats [Chambers and MacAvoy 2000]. Each repeat size represents a class of microsatellites, named di-, tri-, tetra-, penta-, and hexanucleotide, respectively. This has to be indivisible (for example, $-\mathrm{ATAT}-$ is the unit $-\mathrm{AT}-$ repeated two times) and represented by the simplest and "most alphabetical" formula possible, e.g., $-(\mathrm{AC})_{15}-$ rather than $-(\mathrm{TG})_{15}-$ [Chambers and MacAvoy 2000].

There is some disagreement concerning the minimum repeat unit for microsatellites. Some studies do not consider mononucleotides as microsatellites [e.g., Denver et al. 2004], although the general rule of their evolution seems to be the same as for the other classes [Dieringer and Schlötterer 2003]. The maximum size of 6 nt for the repeat unit is also under debate. Studies showed that tandem repeats with larger units (the minisatellites) evolve especially because of errors of recombination, rather than polymerase slippage (see Section 3.2.2 for details). So, there is a window between 6 and 10 nucleotides, where the mechanism of evolution is not certain. These classes are included in some studies as microsatellites [Desmarais et al. 2006].

It is important to note that

$$(\mathrm{AC})_n = (\mathrm{CA})_n = (\mathrm{TG})_n = (\mathrm{GT})_n,$$

while in different reading frames or on a complementary strand, we have

$$(\mathrm{AGC})_n = (\mathrm{GCA})_n = (\mathrm{CAG})_n = (\mathrm{CTG})_n = (\mathrm{TGC})_n = (\mathrm{GCT})_n.$$

Thus, 4 unique classes are possible for dinucleotide repeats, namely, $(\mathrm{AT})_n$, $(\mathrm{AG})_n$, $(\mathrm{AC})_n$, and $(\mathrm{GC})_n$. Similarly, there are 10 unique classes for trinucleotide repeats, 33 for tetranucleotide repeats and so on, and 2 classes are possible for mononucleotide repeats, in the case where the study considers them as microsatellites [Jurka and Pethiyagoda 1995]. However, as the trinucleotide repeats are found often in coding regions of the DNA, the special case of each such locus has to be taken in consideration.

Number of Repeats

Another important parameter for a microsatellite is the number of repeats, but several problems can occur while setting this parameter. The first one involves the total length of the locus. A dinucleotide microsatellite and a hexanucleotide microsatellite, each having 10 repeats, will have different total length of the respective locus. Implicitly, the two loci are exposed to different physical constraints. The hexanucleotide locus, being longer in terms of total number of nucleotides, has more chances to bear mutations (see the Section 3.2.2). The second problem concerns the minimum number of repeats. We can consider as a microsatellite all the elements that contain at least one tandemly repeated given unit. In fact, a size threshold needs to be established, either as a number of repeats [Kruglyak et al. 2000], or as total length of the locus [Toth et al. 2000].

The reasons behind these thresholds are statistical, and no real consensus has been reached on this matter [Ellegren 2004]. Although Rose and Falush [1998] proposed a minimum of eight nucleotide long loci, other studies [see Primmer and Ellegren 1998; Noor et al. 2001] showed that shorter loci are still subjected to polymerase slippage, so they can be included in the microsatellite category. It has to be noticed that these studies have been conducted in different organisms, yeast *Saccharomyces cerevisiae* [Rose and Falush 1998], barn swallow *Hirundo rustica* and pied flycatcher *Ficedula hypoleuca* [Primmer and Ellegren 1998], and different species from the obscura group of Drosophila [Noor et al. 2001]. From a bioinformatic point of view it has been proposed that more than two units have to be present at a locus for this to be considered microsatellite [Delgrange and Rivals 2004]. But these considerations all go back to the basic mechanism of microsatellite formation: a (micro)satellite is generated when the same sequence is inserted repeatedly by the slippage mechanism, which is facilitated by the fact that the presence of repetitive sequences promotes more slippage. It is also possible that microsatellites originate from proto-microsatellites [Wilder and Hollocher 2001].

Proto-microsatellites and quasi-microsatellites

The concept of proto-microsatellites is accepted only if we consider that a repeated sequence needs to reach a certain minimum length for becoming a microsatellite [like Rose and Falush 1998; Chambers and MacAvoy 2000]. In this case proto-microsatellites are defined as repeated sequences with a number of repeats considered too small to be variable. They appear by chance, following point mutations [Jarne and David 1998], within cryptically simple sequences, defined as a scramble of repetitive units lacking a clear tandem arrangement [Hancock 1999].

The quasi-microsatellites are non-repetitive sequences with potential to become repetitive due to several point mutations. For example, the sequence $-\mathrm{ACCT\,ACTT}-$ is a quasi-microsatellite as any of the substitutions $\mathrm{C} \rightarrow \mathrm{T}$ or $\mathrm{T} \rightarrow \mathrm{C}$ results in $-(\mathrm{ACTT})_2-$ or $-(\mathrm{ACCT})_2-$, respectively. A sequence like $-\mathrm{ACC\,ACC\,AGC}-$ is not considered a quasi-microsatellite because it already contains the proto-microsatellite $-(\mathrm{ACC})_2-$. Contrary, the sequence $-\mathrm{ACC\,AGC\,ACC}-$ is a quasi-microsatellite. The number of mutations necessary does not have an exact limit as it depends on the length of the repeat unit and the position of the mutation. Such sequences are frequently found in regions of "cryptic simplicity" [Tautz et al. 1986]. These are regions of variable size, with few repeats of various units, not necessarily adjacent to each other.

Imperfections

By definition, the microsatellites are sequences repeated in tandem subjected to the same molecular constraints as the rest of the genome. They undergo structural changes through addition or removal of repeat units or through point mutations [Eckert and Hile 2009]. The latter event can cause imperfections in these arrays, thus leading to the formation of imperfect microsatellites. Imperfections in the microsatellite tract are thought to interfere with replication slippage by limiting microsatellite expansion [Kruglyak et al. 1998] and their accumulation may lead to the "death" of the microsatellite [Taylor et al. 1999]. The number of imperfections that change a microsatellite locus into a region of "cryptic simplicity" depends on the definition adopted. For example, two microsatellites $-(\mathrm{AC})_{20}-$ separated by three non-repeated bases can be considered as two distinct microsatellites or as a single imperfect one. The majority of the empirical analyses accept imperfections of several bases, while the theoretical studies often consider such loci as two distinct microsatellites even if there is an imperfection of one base [Bell and Jurka

Table 3.1 – Classes of Microsatellites [modified after Chambers and MacAvoy 2000]

Classes	Sequence
Perfect	$-(\mathrm{AC})_n-$
Imperfect (with substitution)	$-(\mathrm{ACT})_n\,\mathrm{CCT}\,(\mathrm{ACT})_m-$
Imperfect (with deletion)	$-(\mathrm{ACT})_n\,\mathrm{CT}\,(\mathrm{ACT})_m-$
Compound	$-(\mathrm{ACT})_n\,(\mathrm{AG})_m-$
Interrupted compound	$-(\mathrm{ACT})_n\,\mathrm{AGAA}\,(\mathrm{AG})_m-$
Complex	$-(\mathrm{ACT})_n\,(\mathrm{AG})_m\,(\mathrm{ACC})_k-$
Interrupted complex	$-(\mathrm{ACT})_n\,(\mathrm{AG})_m\,\mathrm{ATC}\,(\mathrm{ACC})_k-$

1997; Lai and Sun 2003]. Studies concerning only perfect microsatellites do not allow for any imperfection and the sequences studied cannot be in proximity to other repeated sequences of the same unit.

Complexity

The concepts of repeat unit and imperfections mentioned before can be applied only for those microsatellites considered perfect or simple, composed from a unique repeated motiv. Two more types of microsatellites can be defined: compound and complex [Chambers and MacAvoy 2000]. The compound microsatellites are defined as two concatenated microsatellites of distinct repeated units, while the complex microsatellites are a generalized form of compound ones, with more than two distinct units. In both cases the repeat units can be of different size and period, imperfections being possible between them, creating interrupted compound and interrupted complex microsatellites, respectively, and the same units can be repeated within a complex microsatellite (see Table 3.1). Kofler et al. [2008] made one of the few *in silico* surveys of compound microsatellites in eight taxa. These types of microsatellites are often used for studies of population biology. Their evolution hasn't been studied in detail, but they are considered to originate from regions of simple microsatellites that have degenerated [Buschiazzo and Gemmell 2006].

3.1.3 Bioinformatic Detection

Since the existence of large genomic resources, it is possible to isolate *in silico* all the microsatellite loci from a given genome and with this to obtain the exhaustive distribution of all the classes of microsatellites. The genomic sequences have billions of nucleotides, so the extraction of microsatellite data has to be automated by the usage of informatic

Table 3.2 – A Non Exhaustive View upon Studies on Microsatellite Loci Revealing Detection Parameters Inconsistencies

Study	Repeat Unit	Minimum Length	Perfection
Jurka and Pethiyagoda [1995]	1–6	3 repeats and 12 bp	perfect
Bell and Jurka [1997]	2	2 repeats	perfect
Rose and Falush [1998]	1, 2, 4	0 repeats	perfect
Falush and Iwasa [1999]	2	0 repeats	perfect
Toth et al. [2000]	1–6	12 bp	perfect
Katti et al. [2001]	1–4	20–21 bp	imperfect, 1 error/10 bp
Calabrese and Durrett [2003]	2–3	5 repeats	perfect, at least 50 bp appart
Dieringer and Schlötterer [2003]	1–4	2 repeats	perfect
Kayser et al. [2006]	2–5	8 repeats	imperfect
Brandström and Ellegren [2008]	2–5	3 repeats	perfect
Galindo et al. [2009]	1–6	12 bp	perfect
Castagnone-Sereno et al. [2010]	1–6	12, 8, and 5 repeats	perfect
Varela and Amos [2010]	2	2–20 repeats	perfect

algorithms. Given all the studies that have been conducted upon microsatellites based on the genomic sequences alone, one could think it is enough to reutilise the detection methods already used.

A literature search, concerning the algorithms used, reveals several inconsistencies. First, the algorithms and the detection criteria vary amongst studies. This problem concerns mainly the minimum size and the degree of perfection of the loci detected, and this radiates directly from the lack of a formal definition for microsatellite. Examples of such studies can be seen in Table 3.2. Second, most studies do not give details on the algorithm used [e.g., Falush and Iwasa 1999; Lai and Sun 2003] and therefore it does not allow for a clear evaluation of the microsatellites detected. Third, most of the studies evaluate only perfect microsatellites. However, there are several reasons behind this choice: the loci are used for testing theoretical models of microsatellite dynamics that don't integrate imperfections, or the actual detection of imperfect loci is more difficult than perfect ones and it requires more sophisticated algorithms.

There are many published algorithms dedicated to the detection of microsatellites, like `mreps` [Kolpakov 2003], MIcroSAtellite – MISA [Thiel et al. 2003], Sputnik [Abajian 1994], SSRFinder [Gao et al. 2003], Tandem Repeat Finder – TRF [Benson 1999], Tandem Repeat Occurrence Locator – TROLL [Castelo et al. 2002], and many custom programmes [see Sharma et al. 2007, for a rather comprehensive list of the most used ones]. A comparison of the limits of microsatellite detection between some of these programmes has been performend by Leclercq et al. [2007]. In my own research, I used the programme TRF. It scans sequences in order to determine regions where repeat units are periodically

repeated, though not necessarily tandemly repeated, based on a set of statistical rules detailed in the TRF article [Benson 1999]. The most appropriate repeat unit is then determined for each region, and this is aligned with the region using a Wraparound Dynamic Programming algorithm. This procedure takes as input a repeat unit and a sequence, it yields an optimal global alignment between the sequence and a perfect tandem repeat of the respective unit, and it optimizes both the alignment score and the number of repeats of that unit. A score is computed from this alignment by attributing a positive weight to each correctly aligned nucleotide (*matches*), and a negative weight to substitutions (*mismatches*) and to insertions and deletions (*indels*). Alignment weights can be adjusted. When the alignment score is higher than a threshold (that can also be adjusted), the algorithm returns the microsatellite detection with the corresponding consensus repeat unit.

3.1.4 Microsatellite Distribution

It is appropiate to classify microsatellites according to their association with coding sequence, as this is related to the mutational and selective forces that operate on different types of repeats. Microsatellites can be found anywhere in the genome, both in protein-coding and non-coding regions [Arcot et al. 1995; Wilder and Hollocher 2001], but they are not regularly distributed within a genome.

There are $1,600$ eukatyotic genomes available on NCBI (http://www.ncbi.nlm.nih.gov-/genome/browse/). They are not all assembled and the coverage is not optimal, but the sequences of the model organisms are of good quality. The frequency of genomic microsatellites varies per taxon both in terms of numbers and types of repeats [Hancock 1999]. Some of the estimated frequencies for perfect microsatellite loci per genome are 0.85% in *Arabidopsis*, 0.37% in *Zea mays*, 3.21% in *Tetraodon nigroviridis*, 2.12% in *Fugu rubripes* [Roest Crollius et al. 2000], 0.21% in *C. elegans*, and 0.09% in *Meloidogyne incognita* [Castagnone-Sereno et al. 2010], but different scanning techniques yield different results.

Due to the high mutation rate that microsatellite regions experience (see below), it is expected for the coding regions to have a low microsatellite density. This hypothesis has been confirmed by Toth et al. [2000], in a survey on different species. Nonetheless, microsatellites exist in promoter regions and may be sites for protein binding or be near such sites that interfere with gene expression and transcription. In these cases, the number of repeats in the microsatellites has an effect on transcription and the degree

of protein binding. Examples of such loci include those found (a) in association with regulatory regions, like the $-(\mathrm{CT})_n-$ microsatellites in the $5'$-UTR regions of certain *Arabidopsis* genes [see Oliveira et al. 2006, and references therein], (b) located in promoter regions effecting gene activity, like the $-(\mathrm{CT})_n-$ repeat that serves as a transcriptional element for heat-shock protein gene *hsp26* in *Drosophila* [Sandaltzopoulos et al. 1995], (c) located in intronic regions having effects on gene expression regulation, like the $-(\mathrm{TTC})_n-$ element in the first intron of *IIL* gene in *Arabidopsis thaliana* [Sureshkumar et al. 2009], or the $-(\mathrm{CA})_n-$ tract located in the first intron of the EGF receptor gene [Gebhardt et al. 2000]. Irrespective to repeat unit, microsatellites have been found to have roles in the chromosome organization, at least in some plant genomes [Cuadrado and Schwarzacher 1998]. Dimeric repeats seem to play a role in the duplex curvature and supercoiling of the DNA [Baldi and Baisnée 2000] and together with other classes of microsatellites, they are found clustered around centromeric regions [see Li et al. 2002, and references therein]. They also seem to play some role in the regulation of DNA metabolic processes, as many microsatellite and minisatellite regions act as hotspots for recombination [Templeton et al. 2000].

Microsatellites in Nematode Genomes

In my own research article, I made a comparison between the different classes of microsatellite loci in the genomes of *C. elegans* and *P. pacificus*. During my analysis, another study was published that made an analysis on microsatellite loci in a comparative manner between several nematodes: *C. elegans*, *P. pacificus*, *M. incognita*, *M. hapla* and *B. malayi* [Castagnone-Sereno et al. 2010]. Although the results reported for *P. pacificus* by Castagnone-Sereno et al. [2010] differ from the results I have found in my survey, which occured due to the screening procedure, I will resume their findings.

With respect to the abundance of the different classes based on the repeat unit size (i.e., mono to hexanucleotides), the overall distributions varied greatly among the species. Particularly, *B. malayi* exhibited the highest density and genome coverage of microsatellites, except for the hexanucleotides, that were more frequent in the genome of *C. elegans*. The most common repeat size class also varied between the nematodes. For example, AT repeats are more prevalent in *M. hapla*, AG repeats in *P. pacificus* genome, AAAT repeats in *M. incognita*, and ACCT in *C. elegans*. The genome of *C. elegans* has been surveyed for all microsatellite regions in a study by Toth et al. [2000]. They identified the dinucleotide loci to be the most abundent class of microsatellites in the intergenic and

intronic regions, while tri- and hexanucleotide microsatellites are found more often in exonic regions than any of the other classes.

3.2 Mutation Process

The majority of microsatellites in higher organisms are believed to evolve neutrally, meaning that there is no selection pressure on the number of repeats. Nonetheless, some microsatellites do exist in promoter regions and their repeat number has an effect on transcription and the degree of protein binding. The abundance, function and effects of the microsatellites are associated with their mutation rate, which is very high compared with the rates of point mutation at coding loci. The mutation process seems to display distinct differences among species, repeat types, loci and alleles [Brock et al. 1999; Ellegren 2000a; Schlötterer 2000]. The instability is predominantly manifested as changes in the number of repeats by a stepwise mechanism of gain and loss via two mechansims, either DNA slippage during DNA replication [Tachida and Iizuka 1992] or recombination between DNA strands [Harding et al. 1992].

3.2.1 Mutation Models

In order to understand the reason for such high mutation rates for the microsatellite loci, we have to look at the model of mutation itself. A mutation model of their evolution is also necessary if allele frequency data from two groups of individuals are to be used for estimating the genetic distance between them. One of the oldest models for microsatellite evolution is the stepwise mutation model (SMM) originally proposed by Ohta and Kimura for electrophoretic alleles [Ohta and Kimura 1973]. In this model, the number of repeat units is equally likely to increase or decrease by one, at a rate independent of the microsatellite's length. The constraint is that the number of repeat units cannot become smaller than one. This represents a symmetric random walk independent of repeat length [Valdes et al. 1993]. The simplest model of evolution of microsatellites is considered the infinite alleles model (IAM), in which mutations occur at a constant rate and each mutation creates a novel allele [Kimura and Crow 1964]. The SMM model is considered more appropiate for microsatellites.

Numerous additions to this basic model have been introduced. These include (a) allowing the mutation rate to depend on the microsatellite's length [Kruglyak et al. 1998], (b) allow-

ing for larger mutations than of one repeat unit [Di Rienzo et al. 1994], and (c) allowing for point mutations to interrupt microsatellites [Bell and Jurka 1997; Kruglyak et al. 1998]. In the latter two studies mentioned, the authors proposed that there is an equilibrium between point mutations and length mutations. So, the rate of length mutations increases with increasing repeat count, while point mutations break long repeat arrays into smaller units. This can explain why the large microsatellites have a limit of expansion. Still, evolutionary dating of divergence times is not trivial [Calabrese et al. 2001]. The model introduced by Di Rienzo et al. [1994] is a modification of the SMM with most mutations involving a gain or loss of a single repeat, and the remainder of the mutations being multi-step mutations following a geometric distribution.

Evidence from different sources, that have been cited across this chapter, show that the best models have the following properties: (a) long microsatellites are more likely to mutate, (b) long microsatellites have a bias towards contraction and (c) short microsatellites have a bias towards expansion.

3.2.2 Mutation Mechanism

The predominant mechanism by which microsatellites mutate is believed to be the replication slippage [Eisen 1999]. When the DNA replicates, the two strands sometimes disassociate. In nonrepetitive DNA, the strands reassociate in the same way that they were before the disassociation event, with the matching base pairs on the opposing strands. But in repetitive DNA, since there are so many possible matching base pair alignments, sometimes the strands realign differently, forming an unmatched loop on one of the strands. Then, when the two strands completely disassociate and begin replication anew, the strand with the loop will contain a longer microsatellite than the opposing strand. Microsatellites on the template strand will always have the same length before and after the slippage event. If the loop is on the template strand, then the microsatellite on the replicating strand will be shorter, and if the loop is on the replicating strand, then the microsatellite on its side will be longer (Figure 3.1). Most of these primary mutations are corrected by the mismatch repair system, and only the small fraction that was not repaired ends up as microsatellite mutation events [Harr et al. 2002; Tijsterman et al. 2002].

Recombination could potentially change the microsatellite length by unequal crossing-over or by gene conversion [Richard and Pâques 2000]. However, there are substantial

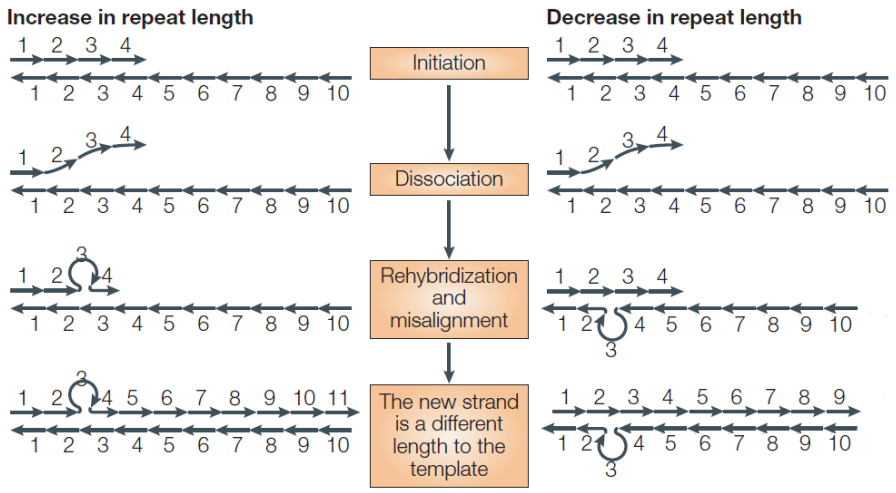


Figure 3.1 – Replication slippage. After the replication of a repeat tract has been initiated, the two strands might dissociate. If the nascent strand then realigns out of register, continued replication will lead to a different length from the template strand. If misalignment introduced a loop on the nascent strand, the end result would be an increase in repeat length. A loop that is formed in the template strand leads to a decrease in repeat length. Figure from Ellegren [2004].

doubts that recombination acts to majorly influence microsatellite variability as it does for minisatellites [Ellegren 2004].

3.2.3 The Life Cycle of the Microsatellites

The concept of *life cycle* (Figure 3.2) for microsatellite loci is a compilation of several ideas about the evolution of microsatellites. Specifically, Messier et al. [1996] and Gordon [1997] noted that «a base substitution allowed the birth of a microsatellite» to debate on how a microsatellite locus appeared during primate evolution. This “birth”, as applied to microsatellites evolution by Messier et al. [1996], refers to a combination of a mutation that produces repeats of sufficient length for slipped-strand mispairing and a second mutation that is caused by the mispairing and produces repeat length variation. Amos [1999] was first to propose a life cycle-like pattern for the evolution microsatellite loci, and later, Taylor et al. [1999] introduced the term “death” because the ultimate result of interruptions and deletion mutations upon microsatellite loci is an unrecognizable homologous DNA sequence left with only a small portion of the original repeat region. Chambers and MacAvoy [2000] completed the concept of life-cycle by suggesting that a dead microsatellite locus could potentially resuscitate.

Despite the fact that it is still under debate, the genesis of microsatellites in genomes

appears to be non-random [Buschiazzo and Gemmell 2006]. Microsatellites are born from regions of "cryptic simplicity"[Tautz et al. 1986] composed of quasi- and proto-microsatellites following point mutations. These microsatellites are called *de novo* as they appear directly from the genomic regions. In contrast, there are so-called "adopted" microsatellites, that are brought into a receptive genomic location via mobile elements [Wilder and Hollocher 2001]. The relative importance of these two ways of microsatellite initiation is not yet completely understood.

The birth of the microsatellite is followed by the expansion phase. As consequence of mutations, the number of repeats grows—it is said that the proto-microsatellite graduated to the mature phase of its life cycle [Buschiazzo and Gemmell 2006]. Some authors debate the existence of a threshold size for initial expansion [Pupko and Graur 1999; Pérez et al. 2005], while others have proposed threshold sizes ranging from 4 to 8 repeats [Rose and Falush 1998; Lai and Sun 2003]. The expansion of a microsatellite depends on the mutation dynamics of the respective locus. Following their initiation, microsatellites are thought to vary in length by gain and loss of repeats via two mechanisms – replication slippage and recombination. Mutation rate in microsatellites is, on average, high and locus dependent, ranging in eukaryotes, from 10^{-7} to 10^{-3} mutations per locus per generation [Primmer et al. 1996; Schug et al. 1997; Vigouroux et al. 2002; Gusmão et al. 2005; Seyfert et al. 2008]. These represent rather static events of a very dynamic picture of mutations observed at allele-, locus-, individual-, and/or taxon-level [Schlötterer et al. 1998; Di Rienzo et al. 1998; Colson and Goldstein 1999; Ellegren 2000b; Shao et al. 2005, and others]. The rate and course of microsatellites mutations are affected by a number of intercorrelated factors: mutation mechanism, nature of microsatellite, genomic context, individual biological context and selective influences [Buschiazzo and Gemmell 2006].

The mutation rate seems to be higher the longer a microsatellite locus is. There is some consensus on the idea that there is an upper length constraint of some sort when microsatellites reach a pending state. This limit is, however, different in various organisms [Estoup et al. 1993; Primmer et al. 1996; Whittaker et al. 2003; Sainudiin et al. 2004]. In fact, microsatellite growth reaches an upper limit when the expansion is hindered by the introduction of imperfections in the repeat array that stabilize the sequence [Kruglyak et al. 1998]. In addition, it is possible that the processes that drive the expansion also duplicate some of the imperfections, leading to their faster accumulation [Estoup et al. 1995; Rolfsmeier and Lahue 2000]. In addition, Harr and Schlötterer [2000] showed that certain mutation mechanisms can delete imperfections, bringing the locus to a smaller, but perfect state.

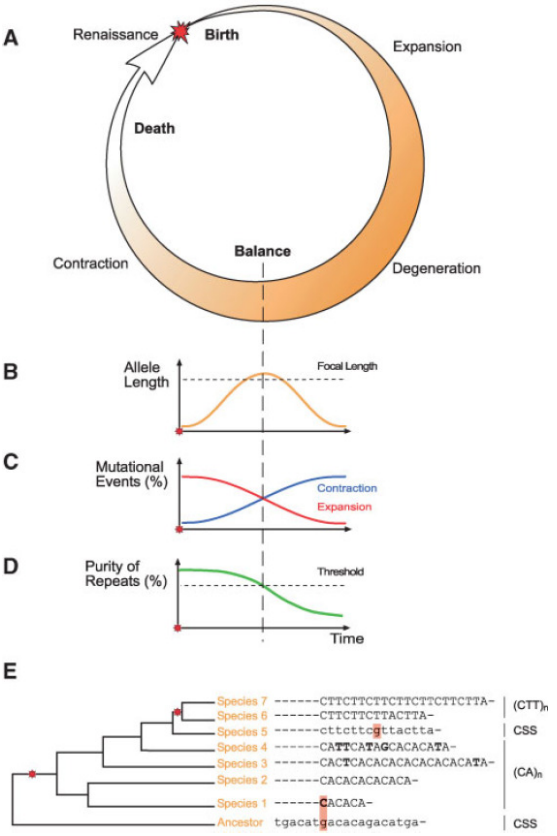


Figure 3.2 – Biology of a Microsatellite Locus from Birth to Resurrection (**A**) Schematic life cycle of a microsatellite locus, (**B**) variation in time of the average allele length, (**C**) relative occurrence of repeat expansion and contraction events over time and (**D**) purity of repeat (proportion of perfect repeats in the array), (**E**) Data superimposed on a phylogeny allow direct observation of a locus life cycle (CSS: Cryptically Simple Sequence) [Buschiazzo and Gemmell 2006].

As consequence of all the mutations, microsatellites pass over the balanced phase of the cycle, where the length of the locus is at maximum, towards the contraction phase. Eventually, the accumulation of interruptions breaks the repeats and leads to a blend of unique DNA sequences that can be considered, again, areas of "cryptic simplicity". The outcome has been termed the "death" of the microsatellite [Taylor et al. 1999]. Its rate seems to be much lower than their birth rate, which explains the enrichment of such loci in the genomes [Buschiazzo and Gemmell 2006]. As regions of "cryptic simplicity" can be the birth places of microsatellites and the death of the loci closes the circle and forms again such regions, these can also be seen as the places of resurrection for microsatellites. [Chambers and MacAvoy 2000]

3.2.4 Mutation Rates and Their Variation

Mutation models are available, but it can be difficult to assess to what extent they reflect true evolutionary processes. The high rate of mutation at microsatellite loci, fortunatelly, makes it possible to observe mutation events directly. There are several types of data sets to consider when studying microsatellites: (a) pedigree data, where it is possible to directly count mutations, (b) *in vitro* experiments, and (c) genome data. Pedigree studies

in humans, reported by Huang et al. [2002], showed that, longer alleles lose repeat units more often than shorter alleles do, but shorter alleles tend to gain repeat units more often than do longer alleles. This conclusion was also the outcome of other studies based on different data types, like Schlötterer et al. [1998] and Harr and Schlötterer [2000] for *D. melanogaster* and Ellegren [2000b] for human (pedigree data). Furthermore, the studies by Schlötterer et al. [1998] estimated the average mutation rate across all loci analysed at 6.3×10^{-6}, in good agreement with another study by Schug et al. [1997], performed on the same number, but different loci and on different *D. melanogaster* lines. From the comparison of these two studies another conclusion can be deduced: mutation rate is locus dependent, because Schug et al. [1997] observed one mutation event while Schlötterer et al. [1998] observed nine mutation events per locus.

In vitro experiments showed that slippage rate increases linearly with microsatellite length and that constrictions are found on a higher rate than expansions [Shinde 2003]. This is, at first glance, little relevant for evolutionary studies, because *Taq* DNA polymerase has no mismatch repair mechanism. But the mismatch repair mechanism can be knocked out for experimental purposes. In study on mismatches-proficient and -deficient *Drosophila* mutation accumulation lines, Harr et al. [2002] showed that the microsatellite instability was highly elevated in those lines in which the mismatch repair-mechanism was deficient. Specifically, the authors assayed dinucleotide loci of the type $(\mathrm{GA})_n$, $(\mathrm{GT})_n$, and $(\mathrm{AT})_n$ and estimated the mutation rate in mismatch repair-deficient lines at 1.43×10^{-3} for short alleles and 1.58×10^{-2} for long alleles. In the wild-type mutation accumulation lines, the mutation rate of a subset of the same loci was at 5.93×10^{-6}, while in the heterozygous lines it was slightly elevated (7.23×10^{-6}), but both were two to three orders of magnitude lower than mutation rates in the mismatch repair-deficient lines.

Genomic data of complete or nearly complete sequenced genomes offers the possibility to have the length distribution of all microsatellites in one idealized individual. Calabrese and Durrett [2003] investigated the dinucleotide repeates in *Drosophila* and human genomes and elaborated on models of microsatellite evolution by considering general slippage processes with uniform point mutation process. The genomic data, together with mutation accumulation experiments offer the possibility of investigating the spontaneous mutations. The analysis of such data offers another possibility to estimate direct mutation rates, and I will summarize some of them in the following paragraphs. One important conclusion, coming from observations of spontaneous mutations, is that the mutation process seems to be heterogeneous with respect to loci, repeat types and organisms. This

makes comparisons between strains and between species extremely difficult. Because the rate of mutations is primarily dependent upon the number of repeats, it is critical to recognize that the evolutionary rate for a single locus will change with the size of the allele. Thus, in an evolutionary tree, a single locus may have rapidly evolving lineages and slowly evolving lineages. Consequently, choosing microsatellite loci with the appropiate rate to address a specific evolutionary frame is difficult if the alleles at a locus span a large size range. The comparison of evolutionary rates and genetic diversity of different species requires a good understanding of the specific number of repeats in each microsatellite allele and the diversity of allele sizes.

Caenorhabditis elegans

To study microsatellite mutation rate in *C. elegans* MA lines, Seyfert et al. [2008] performed a BLASTN search with query sequences of 10 perfect repeating units for each unique dimer, trimer and tetramer. The largest two loci idetified for each trimer and tetramer motif were selected for analysis, while the dimers ranged between 14 and 191 repeat units. In total, 29 loci with a large variety of sizes were analysed in a set of 80 MA lines at 140^{th} generation. All the dimer loci assayed had the repeat unit, $-\mathrm{GT}-$, and, according to the analysis, their mutation rate increased exponentially with repeat size from, 9.0×10^{-5} to 1.8×10^{-2} per allele per generation. The same correlation was observed for trimer and tetramer loci, while for several small loci no mutations were observed. From all 273 mutations, 73% involved changes of more than one repeat. Large insertions were observed at only one locus, while large deletions occured in dimer loci of more that 50 repeat untis in length, but not at the largest loci. The microsatellite loci in *C. elegans* seem to have a tendency towards lengthening, as there are much more addition than deletion events, but this balance is reversed if two particular loci are excluded from the analysis. It is important to note that the microsatellites analysed do not appear to follow a simple stepwise mutation model. Many of the observed mutations may represent changes of more that one unit at a time, rather than accumulated single step mutations.

Daphnia pulex

Seyfert et al. [2008] evaluated the dinucleote microsatellite mutation rate in 268 mutation accumulation lines of *D. pulex* after 27 generations. The average mutation rate was 1.16×10^{-4} per allele per generation for long loci, an order of magnitude higher that the

average for the short loci (< 30 repeats or less), which was 7.11×10^{-5} per allele per generation. Most of the mutations were additions of more than a single repeat. These average rates are similar to the mutation rates found for the *C. elegans* MA lines.

Arabidopsis thaliana

Marriage et al. [2009] estimated the rate of mutation per allele per generation for 54 dinucleotide repeats in 96 *A. thaliana* MA lines at the 30^{th} generation. The average estimated rate across all loci was 8.87×10^{-4} and the majority of mutations were gains or losses of a single repeat. The mutation rate was heterogeneous across loci and increased with repeat number. Mutations at longer alleles were more frequently losses than gains. The authors observed a strong dependency of the mutation rate on repeat unit and initial length of the locus, suggesting that the average genomic mutation rate depends on the relative frequency of the various repeat types and on the distribution of allele sizes segregating in the population.

3.3 Instead of Conclusion

Much of the biology of microsatellites is well described, but only partially understood, and the evolutionary processes of simple repeats is far from simple. The genetic markers depend crucially on the underlying mutation processes that generate variation. And these are still poorly understood. The mutation mechanisms can explain the behaviour of microsatellite regions only partially. The mutation rates depend on so many factors…

3.3.1 Reasons for This Chapter

or Why This Is Directly Relevant to My Line of Research

In this chapter, I made a rather comprehensive overview of the biology of microsatellites and their mutation process. It is important to understand the genome biology of these elements in order to attempt to characterize them in a given species. The definition and molecular structure of a microsatellite, as shown in Section 3.1, has a great impact on the algorithms and parameters used during genome mining. Understanding the mutation process of these sequences is complicated, as they are under several evolution constraints that depend on many factors. In Section 3.2, I have presented a comprehensive overview

of this process and I have reviewed several studies that help explain how and whhy the mutation rate is so heterogenenous.

In my research article, I am describing the microsatellite content of the genomes of *P. pacificus* and I am estimating the mutation rate of several loci randomly chosen across the entire genome, with the help of the mutation accumulation line experimental setup. This work will be described in detail in Chapter 5.

Aim of the Thesis

The rates and properties of spontaneous mutations are important for many questions in evolutionary biology. Under neutrality, the rate of molecular evolution is expected to be equal to the mutation rate, so between-species molecular divergence can be used to date divergence times of species by assuming clock-like molecular evolution. Conversely, the rate of molecular divergence at silent sites between species can be used to estimate the mutation rate. All these require the assumption of neutrality, values for the generation time and divergence dates of species. Mutation accumulation experiments are an easy way to study new mutational variation. Using this setup it is possible to directly assess the mutation rates at various regions in the genomes and calculate the divergence between strains or species. Self-fertilizing organisms are well-suited for such experiments as it is possible to create a bottleneck of one individual each generation for multiple generations. The nematode *Pristionchus pacificus* (Family: Diplogastridae) is the organism of choice for this thesis as it has been developed as a model organism in various comparative studies to *Caenorhabditis elegans* (Family: Rhabditidae).

The aim of this thesis is to study evolutionary processes at the level of the mitochondrial DNA and microsatellite loci. These two regions are often assumed under full neutrality, being, therefore, suited for mutation rate and divergence estimations. With the help of mutation accumulation lines and several wild strains of *P. pacificus*, I adressed the following:

1. What is the mutation rate in the mitochondrial DNA of *P. pacificus*?
   and, related to this,
2. What is the divergence time between wild isolates of *P. pacificus*?
   and,

3. How does the time to the most recent common ancestor correlate with the ecology of *P. pacificus*?
4. What is the composition and distribution of the microsatellite regions in the genome of the nematode *P. pacificus*?
   and,
5. What are the mutation rates of the different microsatellites of the *P. pacificus* nuclear genome?

Part II

Results and Discussion

Chapter 4

Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*

Ruxandra I. Molnar, Gabi Bartelmes, Iris Dinkelacker, Hanh Witte, and Ralf J. Sommer
Molecular Biology and Evolution, 28: 2317-2326, 2011.

4.1 Synopsis

This study describes the sequence and organization of the mitochondrial genome of *P. pacificus* and the mutation spectrum and rates found in the mutation accumulation line experiment. Using these mutation rates, we estimated the time to the most recent common ancestor for nine *P. pacificus* strains, isolated worldwide.

The mitochondrial genome of *P. pacificus* is a circular molecule of $15,954$ bp in length, well into the size limits of the nematode mtDNAs sequenced to date. Comparing to other nematodes, the *P. pacificus* mtDNA has two new features: (i) an intergenic region without any sequence similarity to any intergenic regions of other nematodes, which is situated between tRNA-Pro and tRNA-Val, and (ii) a predicted suppressor-tRNA for the codon UAA, situated in the D-loop, that we think influenced the outcome of the mutations found.

One hundred MA lines were initiated from the F3 descendents of a single, inbred, wild-

type strain of *P. pacificus*. They were propagated for a period of 142 generations, from a single, randomly selected J3 larva from the middle of the parental reproductive period. From the original 100 lines, 82 survived until the 142^{nd} generation, resulting in a survival rate of 87.3% per 100 generations.

The entire mtDNA, with the exception of the D-loop, from each MA line was sequenced and assembled independently. The subsequent calculations are all based on $13,500$ bp in each of the 82 MA lines at the 142^{nd} generation. Twelve mutations have been observed, to a total mutation rate $\mu_{\text{total}} = 7.6\,(\pm 2.2) \times 10^{-8}$ per site per generation. Seven of these mutations were base substitutions, yielding the mutation rate for base substitutions at $\mu_{\text{bs}} = 4.5 \times 10^{-8}$. Five indels were found in the MA lines, four deletions occured at T-homopolymeric loci in four different genes and lines, and one insertion at the level of a non-repetitive sequence. Interestingly, one of the base substitution mutations and all the indels resulted in a direct premature STOP codon, and, respectively, 2–3 amino acids downstream of the mutation. This degree of premature STOP codons with no drastic effect on the fitness of the MA lines has never been observed until now in other similar studies.

Time estimates are essential in phylogeographic interpretations of population subdivisions, expansions, migrations, and colonization of new areas. Given that recent studies on *P. pacificus* focused on island invasion in one of the biodiversity hotspots, namely the La Réunion Island, a minimal divergence time between some of the strains collected there was essential. We used the mutation rate estimates obtained from the MA line experiment to calculate the divergence between several strains of *P. pacificus*. These are the first divergence estimates between different strains of nematodes reported in literature. The closest strain to our laboratory wildtype strain, *P. pacificus* California, is one of the strains from La Réunion island at 2.9×10^5 generations, while the most recent split occured 2.3×10^4 generations between the strain from Madagascar and another strain from La Réunion. All three nested models used to estimate the time to the most recent common ancestor suggested a 10^5–10^6 generations divergence. In laboratory conditions, *P. pacificus* reaches about 100 generations per year. In contrast, one generation per year represents the absolute minimum observed after studying the survival rate of dauer larvae. Furthermore, in nature, *P. pacificus* nematodes are found in association with scarab beetles, that are known to have slow life-cycles. Assuming there are between one and ten generations per year in nature, the divergence between the strains analysed is of 10^4–10^5 years. We prefer, however, to express the divergence times in generations and not in years.

4.2 Contributions

All experiments and analyses were carried out by myself, except for the propagation of the mutation accumulation lines, done by Iris Dinkelacker and Hanh Witte. Gabi Bartelmes helped with the PCR amplification and sequencing of some of the MA lines mtDNA. Experiments planning and manuscript preparation were done by myself and my supervisor Ralf J. Sommer. In total, my contribution to this work was about 70%.

Chapter 5

Tandem repeat patterns and mutation rates in microsatellites of the nematode model organism *Pristionchus pacificus*

Ruxandra I. Molnar, Hanh Witte, Iris Dinkelacker, Laure Villate and Ralf J. Sommer
G3: Genes, Genomes, Genetics, in review.

5.1 Synopsis

This study describes the microsatellite loci present in the genome of *P. pacificus*, and the mutation spectrum and rates found in the mutation accumulation experiment. We used the same set of MA lines as in the mitochondrial study.

The program Tandem Repeats Finder was used to detect the presence of tandemly repeated nucleotides, and two data sets were generated. They differ in number of loci found, due to the stringency parameters. Specifically, in the TRF-loose data set comprised $70,543$ perfect loci, while TRF-strict method contained only 730 loci. For the latter data set, in detail analysis showed that, in general, CG-rich loci are very few or completely missing, when compared with AT-rich loci. As the AT content of the *P. pacificus* genome is about 58%, the theory of random appearance of microsatellite loci from simple sequences predicts to find the AT repeats to be the most common, followed by AC and AG repeats in approximate equal numbers, and only few CG loci. This pattern is not followed, however, by the dimer repeat composition found in the *P. pacificus* genome, arguing against the

expectation of cryptic simplicity. In contrast, the tri- to hexanucleotide repeat loci do follow this expectation, with AT-rich repeats being more abundant than the others.

We estimated the mutation rates of 41 microsatellite loci, with repeat unit sizes of di- to hexanucleotides, randomly chosen in the genome. Their mutation rate ranges between 8.9×10^{-5} and 7.5×10^{-4} per allele per generation. Although the loci have been chosen randomly, eight out of ten dinucleotide loci are of the type $(\mathrm{AG})_n$. This correlates with the general composition of the genome of *P. pacificus*, that has more AG repeats than other dinucleotide repeats. The mutation patterns observed in *P. pacificus* support several features common in the theory of microsatellite evolution: (a) large loci have, on average, more mutations than the small ones, (b) imperfect loci are less prone to accumulate mutations than perfect ones, (c) the mutation process is upwardly biased, as loci tend to gain units more frequently than they lose units.

Our analysis of the *P. pacificus* tandem repeat pattern and mutation rates of microsatellite loci help elucidate the evolutionary forces acting on nematode genomes. The rate and pattern of mutations observed in the MA lines have implications in the use of microsatellites for inference of genetic history. It is critical to recognize that the evolutionary rate for a single locus will change with the size of the allele. Thus, choosing a microsatellite locus with the appropriate evolutionary rate to address a specific evolutionary time frame requires a careful consideration of allele size ranges. The data presented in this study provides a catalog for the selection of adequate markers for studying recent and ancient evolutionary branches of *P. pacificus*.

5.2 Contributions

I performend all the experiments and analyses carried out in this paper. The propagation of the mutation accumulation lines was done by Hanh Witte and Iris Dinkelacker. Laure Villate helped with marker design and microsatellite scoring. Experiments were planned by myself, Laure Villate and Ralf J. Sommer. The manuscript was prepared by myself and Ralf J. Sommer. In total, my contribution to this work was about 90%.

Curriculum Vitae

Molnar I. Ruxandra

24^{th} of March, 1981
Bucharest, Romania

Max Planck Institute for Developmental Biology, Tübingen, Germany || **2005–2011**
PhD Student in the Department of Evolutionary Biology, Prof. Dr. Ralf J. Sommer
Thesis title: "Molecular Evolution and Mutation Accumulation Lines in the Nematode *Pristionchus pacificus*"

R and Bioconductor for Genomic Analysis, Český Krumlov, Czech Republic || **2012**
Workshop on Comparative Genomics, Český Krumlov, Czech Republic || **2011**
Workshop on Molecular Evolution, Český Krumlov, Czech Republic || **2010**

University of Bucharest, Romania || **1999–2005**
1999–2003: Faculty of Biology, Department of Biochemistry
2002–2003: Internship in the Laboratory of Lipoproteins and Atherogenesis, Institute of Cellular Biology and Pathology "Nicolae Simionescu", Bucharest.
Diploma thesis title: "Study for Native and Modified LDL Receptors on Macrophage Cell Cultures"
2003–2005: Master of Biochemistry and Molecular Biology
April 2005–September 2005: Erasmus Fellowship for the master thesis experiments at Kekulé-Institute, Department of Biochemistry, University of Bonn, Germany
Master thesis title: "The Effect of UVA Radiation upon the Cerebellum Neurons in Presence or Absence of Ectoins"

Bilingual High School "George Coşbuc", Bucharest, Romania || **1995–1999**
Specialization: Mathematics and Physics

Bibliography

Abajian C. Sputnik - http://espressosoftware.com/sputnik/, 1994.

Abouelhoda M, El-Kalioby M, and Giegerich R. WAMI: a web server for the analysis of minisatellite maps. *BMC Evolutionary Biology*, 10:167–176, 2010.

Al Rawi S, Louvet-Vallée S, Djeddi A, Sachse M, Culetto E, Hajjar C, Boyd L, Legouis R, and Galy V. Postfertilization Autophagy of Sperm Organelles Prevents Paternal Mitochondrial DNA Transmission. *Science*, 334:1144–1147, 2011.

Allen JF. Why chloroplasts and mitochondria contain genomes. *Comparative and Functional Genomics*, 4: 31–36, 2003.

Amos W. A comparative approach to study the evolution of microsatellites. In Schlötterer C and Goldstein D, editors, *Microsatellites - Evolution and Applications*, pages 60–79. Oxford University Press, New York, 1999.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, and Young IG. Sequence and organization of the human mitochondrial genome. *Nature*, 290:457–465, 1981.

Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UCM, Podowski RM, Naeslund AK, Eriksson A-S, Winkler HH, and Kurland CG. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396:133–143, 1998.

Arcot SS, Wang Z, Weber JL, Deininger PL, and Batzer MA. Alu repeats: a source for the genesis of primate microsatellites. *Genomics*, 29:136–144, 1995.

Armour JA and Jeffreys AJ. Biology and applications of human minisatellite loci. *Current Opinion in Genetics and Development*, 2:850–856, 1992.

Avise JC. The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, 7:371–379, 1998.

Awadalla P, Eyre-Walker A, and Maynard Smith J. Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA. *Science*, 286:2524–2525, 1999.

Baldi P and Baisnée P-F. Sequence analysis by additive scales: DNA structure for sequences and repeats of all lengths. *Bioinformatics*, 16:865–889, 2000.

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

Barrière A and Félix M-A. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Current Biology*, 15:1176–1184, 2005.

Bass AL, Good DA, Bjorndal KA, Richardson JI, Hillis ZM, Horrocks JA, and Bowen BW. Testing models of female reproductive migratory behaviour and population structure in the Caribbean hawksbill turtle, *Eretmochelys imbricata*, with mtDNA sequences. *Molecular Ecology*, 5:321–328, 1996.

Bazin E, Glémin S, and Galtier N. Population size does not influence mitochondrial genetic diversity in animals. *Science*, 312:570–572, 2006.

Beagley CT, Okimoto R, and Wolstenholme DR. The Mitochondrial Genome of the Sea Anemone *Metridium senile* (Cnidaria): Introns, a Paucity of tRNA Genes, and a Near-Standard Genetic Code. *Genetics*, 148: 1091–1108, 1998.

Bell GI and Jurka J. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *Journal of Molecular Evolution*, 44:414–421, 1997.

Bendich AJ. Circular chloroplast chromosomes: the grand illusion. *The Plant Cell*, 16:1661–1666, 2004.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27: 573–580, 1999.

Bento G, Ogawa A, and Sommer RJ. Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution. *Nature*, 466:494–497, 2010.

Berg OG and Kurland CG. Why mitochondrial genes are most often found in nuclei. *Molecular Biology and Evolution*, 17:951–961, 2000.

Bininda-Emonds ORP. Fast Genes and Slow Clades: Comparative Rates of Molecular Evolution in Mammals. *Evolutionary Bioinformatics*, 3:59–85, 2007.

Birky Jr. CW. Uniparental Inheritance of Mitochondrial and Chloroplast Genes: Mechanisms and Evolution. *Proceedings of the National Academy of Sciences*, 92:11331–11338, 1995.

Birky Jr. CW. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics*, 35:125–148, 2001.

Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, and Thomas WK. A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392:71–75, 1998.

Blouin MS. Neutrality Tests on mtDNA: Unusual Results from Nematodes. *The Journal of Heredity*, 91: 146–176, 2000.

Borchert N, Dieterich C, Krug K, Schütz W, Jung S, Nordheim A, Sommer RJ, and Macek B. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Research*, 20: 837–846, 2010.

Brandström M and Ellegren H. Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Research*, 18:881–887, 2008.

Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*, 77:71–94, 1974.

Brenner S. In the beginning was the worm... *Genetics*, 182:413–415, 2009.

Brock GJ, Anderson NH, and Monckton DG. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Human Molecular Genetics*, 8:1061–1067, 1999.

Brown WM, George M, and Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76:1967–1971, 1979.

Burger G, Lang BF, Reith M, and Gray MW. Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proceedings of the National Academy of Sciences*, 93:2328–2332, 1996.

Burger G, Gray MW, and Lang BF. Mitochondrial genomes: anything goes. *Trends in Genetics*, 19:709–716, 2003.

Burzynski A, Zbawicka M, Skibinski DOF, and Wenne R. Evidence for Recombination of mtDNA in the Marine Mussel *Mytilus* trossulus from the Baltic. *Molecular Biology and Evolution*, 20:388–392, 2003.

Buschiazzo E and Gemmell NJ. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*, 28:1040–1050, 2006.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, and Hartl Daniel L. The cost of inbreeding in Arabidopsis. *Nature*, 416:531–534, 2002.

Calabrese P and Durrett R. Dinucleotide repeats in the *Drosophila* and human genomes have complex, length-dependent mutation processes. *Molecular Biology and Evolution*, 20:715–725, 2003.

Calabrese PP, Durrett RT, and Aquadro CF. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics*, 159:839–52, 2001.

Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF, and Johnson KP. Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). *BMC Genomics*, 12:394, 2011.

Cann RL, Stoneking M, and Wilson AC. Mitochondrial DNA and human evolution. *Nature*, 325:31–36, 1987.

Castagnone-Sereno P, Danchin EGJ, Deleury E, Guillemaud T, Malausa T, and Abad P. Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne* incognita. *BMC Genomics*, 11:598, 2010.

Castelo AT, Martins W, and Gao GR. TROLL - Tandem Repeat Occurrence Locator. *Bioinformatics*, 18: 634–636, 2002.

Castoe TA, Jiang ZJ, Gu W, Wang ZO, and Pollock DD. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE*, 3:e2201, 2008.

Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, and Pollock DD. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences*, 106:8986–8991, 2009.

Cavalier-Smith T. The origin of nuclei and of eukaryotic cells. *Nature*, 256:463–468, 1975.

Chalfie M, Horvitz HR, and Sulston JE. Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell*, 24:59–69, 1981.

Chambers GK and MacAvoy ES. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B*, 126:455–476, 2000.

Charlesworth B, Sniegowski PD, and Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371:215–220, 1994.

Charlesworth D and Wright SI. Breeding systems and genome evolution. *Current Opinion in Genetics and Development*, 11:685–90, 2001.

Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl H-HM, Turnbull DM, Lightowlers RN, and Howell Neil. The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends in Genetics*, 16:500–505, 2000.

Clayton DA. Mitochondrial DNA Replication: What We KNow. *IUBMB Life*, 55:213–217, 2003.

Coghlan A and Wolfe KH. Four-fold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Research*, 12:857–867, 2002.

Colson I and Goldstein DB. Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics*, 152:617–627, 1999.

Conway DJ, Fanello C, Lloyd JM, Al-Joubori BM, Baloch AH, Somanath SD, Roper C, Oduola AM, Mulder B, Povoa MM, Singh B, and Thomas AW. Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA. *Molecular and Biochemical Parasitology*, 111:163–171, 2000.

Cuadrado A and Schwarzacher T. The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma*, 107:587–594, 1998.

Dawid IB and Blackler AW. Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. *Developmental Biology*, 29:152–161, 1972.

de Bruijn MHL. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature*, 304:234–241, 1983.

Delgrange O and Rivals E. STAR: an algorithm to Search for Tandem Approximate Repeats. *Bioinformatics*, 20:2812–2820, 2004.

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, and Schierwater B. Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proceedings of the National Academy of Sciences*, 103:8751–8756, 2006.

Denver DR, Morris K, Lynch M, Vassilieva LL, and Thomas WK. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*, 289:2342–2344, 2000.

Denver DR, Morris K, Kewalramani A, Harris KE, Chow A, Estes S, Lynch M, and Thomas WK. Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *Journal of Molecular Evolution*, 58:584–595, 2004.

Desmarais E, Belkhir K, Garza JC, and Bonhomme F. Local Mutagenic Impact of Insertions of LTR Retrotransposons on the Mouse Genome. *Journal of Molecular Evolution*, 63:662–675, 2006.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, and Freimer NB. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, 91: 3166–3170, 1994.

Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, and Barch DH. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics*, 148:1269–1284, 1998.

Dieringer D and Schlötterer C. Two Distinct Modes of Microsatellite Mutation Processes : Evidence From the Complete Genomic Sequences of Nine Species. *Genome Research*, 13:2242–2251, 2003.

Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, Mitreva M, Roeseler W, Tian H, Witte H, Yang SP, Wilson RK, and Sommer RJ. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nature Genetics*, 40:1193–1198, 2008.

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

Duchêne A-M, Pujol C, and Maréchal-Drouard L. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current Genetics*, 55:1–18, 2009.

Eckert KA and Hile SE. Every Microsatellite is Different: Intrinsic DNA Features Dictate Mutagenesis of Common Microsatellites Present in the Human Genome. *Molecular Carcinogenesis*, 48:379–388, 2009.

Edwards A, Civitello A, Hammond HA, and Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, 49:746–756, 1991.

Eisen JA. Mechanistic basis for microsatellite instability. In Goldstein DB and Schlötterer C, editors, *Microsatellites Evolution and Applications*, pages 34–48. Oxford University Press, 1999.

Eisenmann DM, Maloof JN, Simske JS, Kenyon C, and Kim SK. The β-catenin homolog BAR-1 and LET-60 Ras coordinately regulate the Hox gene *lin-39* during *Caenorhabditis elegans* vulval development. *Development*, 3680:3667–3680, 1998.

Eizinger A and Sommer RJ. Nematode Epidermal Cell Fates The Homeotic Gene *lin-39* and the Evolution of Nematode Epidermal Cell Fates. *Science*, 452:452–455, 1997.

Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, 16:551–558, 2000a.

Ellegren H. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, 24:400–402, 2000b.

Ellegren H. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5:435–445, 2004.

Estoup A, Solignac M, Harry M, and Cornuet JM. Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Research*, 21:1427–1431, 1993.

Estoup A, Garnery L, Solignac M, and Cornuet J-M. Microsatellite Variation in Honey Bee (*Apis mellifera* L.) Populations: Hierarchical Genetic Structure and Test of the Infinite Allele and Stepwise Mutation Models. *Genetics*, 140:679–695, 1995.

Eyre-Walker A, Smith NH, and Maynard Smith J. How clonal are human mitochondria? *Proceedings of the Royal Society Biology Section*, 266:477–483, 1999a.

Eyre-Walker A, Smith NH, and Maynard Smith J. Reply to Macaulay et al. (1999): mitochondrial DNA recombination - reasons to panic. *Proceedings of the Royal Society Biology Section*, 266:2041–2042, 1999b.

Falush D and Iwasa Y. Size-Dependent Mutability and Microsatellite Constraints. *Molecular Biology and Evolution*, 16:960–966, 1999.

Félix M-A, De Ley P, Sommer RJ, Frisse L, Nadler SA, Thomas WK, Vanfleteren J, and Sternberg PW. Evolution of vulva development in the Cephalobina (Nematoda). *Developmental Biology*, 221:68–86, 2000.

Fontaine KM, Cooley JR, and Simon C. Evidence for paternal leakage in hybrid periodical cicadas (Hemiptera: Magicicada spp.). *PLoS ONE*, 2:e892, 2007.

Forget L, Ustinova J, Wang Z, Huss VAR, and Lang BF. *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Molecular Biology and Evolution*, 19: 310–319, 2002.

Fox TD. Natural variation in the genetic code. *Annual Review of Genetics*, 21:67–91, 1987.

Galindo CL, McIver LJ, McCormick JF, Skinner MA, Xie Y, Gelhausen RA, Ng K, Kumar NM, and Garner HR. Global microsatellite content distinguishes humans, primates, animals, and plants. *Molecular Biology and Evolution*, 26:2809–2819, 2009.

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

Galtier N, Nabholz B, Glémin S, and Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18:4541–4550, 2009.

Gao L, Tang J, Li H, and Jia J. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding*, 12:245–261, 2003.

Gebhardt F, Bürger H, and Brandt B. Modulation of EGFR gene transcription by secondary structures, a polymorphic repetitive sequence and mutations - a link between genetics and epigenetics. *Histology and Histopathology*, 15:929–936, 2000.

Gerber AS, Loggins R, Kumar S, and Dowling TE. Does Nonneutral Evolution Shape Observed Patterns of DNA Variation in Animal Mitochondrial Genomes? *Annual Review of Genetics*, 35:539–566, 2001.

Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Allen JE, Delcher AL, Guiliano DB, Miranda-Saavedra D, Angiuoli SV, Creasy T, Amedeo P, Haas B, El-Sayed NM, Wortman JR, Feldblyum T, Tallon L, Schatz M, Shumway M, Koo H, Salzberg SL, Schobel S, Pertea M, Pop M, White O, Barton GJ, Carlow CKS, Crawford MJ, Daub J, Dimmic MW, Estes CF, Foster JM, Ganatra M, Gregory WF, Johnson NM, Jin J, Komuniecki R, Korf I, Kumar S, Laney S, Li B-W, Li W, Lindblom TH, Lustigman S, Ma D, Maina CV, Martin DMA, McCarter JP, McReynolds L, Mitreva M, Nutman TB, Parkinson J, Peregrín-Alvarez JM, Poole C, Ren Q, Saunders L, Sluder AE, Smith K, Stanke M, Unnasch TR, Ware J, Wei AD, Weil G, Williams DJ, Zhang Y, Williams SA, Fraser-Liggett C, Slatko B, Blaxter ML, and Scott AL. Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science*, 317:1756–1760, 2009.

Gibson T, Blok VC, Phillips MS, Hong G, Kumarasinghe D, Riley IT, and Dowton M. The mitochondrial subgenomes of the nematode *Globodera pallida* are mosaics: evidence of recombination in an animal mitochondrial genome. *Journal of Molecular Evolution*, 64:463–471, 2007.

Gillespie JH. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics*, 155:909–919, 2000.

Gissi C, Reyes A, Pesole G, and Saccone C. Lineage-specific evolutionary rate in mammalian mtDNA. *Molecular Biology and Evolution*, 17:1022–1031, 2000.

Gissi C, Iannelli F, and Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101:301–320, 2008.

Glémin S, Bazin E, and Charlesworth D. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society Biology Section*, 273:3011–3019, 2006.

Gordon AJ. Microsatellite birth register. *Journal of Molecular Evolution*, 45:337–338, 1997.

Gray MW. The Endosymbiont Hypothesis Revisited. In Wolstenholme DR and Jeon KW, editors, *International Review of Cytology - A Survey of Cell Biology*, pages 233–357. Academic Press, 1992. ISBN 0123645441.

Gray MW, Burger G, and Lang BF. Mitochondrial evolution. *Science*, 283:1476–1481, 1999.

Green JL, Inoue T, and Sternberg PW. Opposing Wnt pathways orient cell polarity during organogenesis. *Cell*, 134(4):646–656, 2008.

Grossman LI, Wildman DE, Schmidt TR, and Goodman M. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends in Genetics*, 20:578–585, 2004.

Gusmão L, Sánchez-Diz P, Calafell F, Martín P, Alonso CA, Alvarez-Fernández F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfán MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Góes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A, and Amorim A. Mutation rates at Y chromosome specific microsatellites. *Human Mutation*, 26:520–528, 2005.

Gutierrez A and Sommer RJ. Functional diversification of the nematode mbd2/3 gene between *Pristionchus pacificus* and *Caenorhabditis elegans*. *BMC Genetics*, 8:57, 2007.

Gyllensten U, Wharton D, Josefsson A, and Wilson AC. Paternal inheritance of mitochondrial DNA in mice. *Nature*, 352:255–257, 1991.

Haag-Liautard C, Coffey Nicole, Houle D, Lynch M, Charlesworth B, and Keightley PD. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology*, 6:e204, 2008.

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, and Bowden DK. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proceedings of the Royal Society Biology Section*, 266:485–492, 1999.

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, and Bowden DK. Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. *Proceedings*

of the Royal Society Biology Section, 267:1595–1596, 2000.

Halligan Daniel L. and Keightley PD. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40:151–172, 2009.

Hancock JM. Microsatellites and other simple sequences: genomic context and mutational mechanism. In Goldstein D and Schlötterer C, editors, *Microsatellites - Evolution and Applications*, pages 1–9. Oxford University Press, New York, 1999.

Harding RM, Boyce AJ, and Clegg JB. The evolution of tandemly repetitive DNA: recombination rules. *Genetics*, 132:847–859, 1992.

Harr B and Schlötterer C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics*, 155: 1213–20, 2000.

Harr B, Todorova J, and Schlötterer C. Mismatch repair-driven mutational bias in *D. melanogaster*. *Molecular Cell*, 10:199–205, 2002.

Haynes S, Mozer B, Bhatiadey N, and Dawid I. The locus, a maternal effect homeotic gene, encodes apparent membrane proteins. *Developmental Biology*, 134:246–257, 1989.

He Y, Jones J, Armstrong M, Lamberti F, and Moens M. The mitochondrial genome of *Xiphinema americanum sensu stricto* (Nematoda: Enoplea): considerable economization in the length and structural features of encoded genes. *Journal of Molecular Evolution*, 61:819–833, 2005.

Hebert PDN, Cywinska A, Ball SL, and DeWaard JR. Biological identifications through DNA barcodes. *Proceedings of the Royal Society Biology Section*, 270:313–321, 2003.

Herrmann M, Mayer WE, and Sommer RJ. Nematodes of the genus *Pristionchus* are closely associated with scarab beetles and the Colorado potato beetle in Western Europe. *Zoology*, 109:96–108, 2006a.

Herrmann M, Mayer WE, and Sommer RJ. Sex, bugs and Haldane's rule: the nematode genus *Pristionchus* in the United States. *Frontiers in Zoology*, 3:14, 2006b.

Herrmann M, Mayer WE, Hong RL, Kienle S, Minasaki R, and Sommer RJ. The nematode *Pristionchus pacificus* (Nematoda: Diplogastridae) is associated with the oriental beetle *Exomala orientalis* (Coleoptera: Scarabaeidae) in Japan. *Zoological Science*, 24:883–9, 2007.

Herrmann M, Kienle S, Rochat J, Mayer WE, and Sommer RJ. Haplotype diversity of the nematode *Pristionchus pacificus* on Réunion in the Indian Ocean suggests multiple independent invasions. *Biological Journal of the Linnean Society*, 100:170–179, 2010.

Hoarau G, Holla S, Lescasse R, Stam WT, and Olsen JL. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, 19: 2261–2264, 2002.

Holterman M, van der Wurff A, van den Elsen S, van Megen H, Bongers T, Holovachov O, Bakker J, and Helder J. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Molecular Biology and Evolution*, 23:1792–1800, 2006.

Hong RL and Sommer RJ. *Pristionchus pacificus*: a well-rounded nematode. *BioEssays*, 28:651–659, 2006.

Hong RL, Svatos A, Herrmann M, and Sommer RJ. Species-specific recognition of beetle cues by the

nematode *Pristionchus maupasi*. *Evolution and Development*, 279:273–279, 2008.

Houle D and Nuzhdin Sergey V. Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. *Genetical Research*, 83:7–18, 2004.

Howe DK, Baer CF, and Denver DR. High rate of large deletions in *Caenorhabditis briggsae* mitochondrial genome mutation processes. *Genome Biology and Evolution*, 2:29–38, 2010.

Hu M, Chilton NB, and Gasser RB. The Mitochondrial Genomes of the Human Hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *International Journal for Parasitology*, 32: 145–158, 2002.

Hu M, Chilton NB, and Gasser RB. The Mitochondrial Genome of *Strongyloides stercoralis* (Nematoda)–Idiosyncratic Gene Order and Evolutionary Implications. *International Journal for Parasitology*, 33: 1393–1408, 2003.

Huang Q-Y, Xu F-H, Shen H, Deng H-Y, Liu Y-J, Liu Y-Z, Li J-L, Recker RR, and Deng H-W. Mutation patterns at dinucleotide microsatellite loci in humans. *American Journal of Human Genetics*, 70:625–634, 2002.

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, and Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics*, 12:417–420, 1996.

Jacob JEM, Vanholme B, Van Leeuwen T, and Gheysen G. A unique genetic code change in the mitochondrial genome of the parasitic nematode Radopholus similis. *BMC Research Notes*, 2:192, 2009.

Jannotti-Passos LK, Souza CP, Parra JC, and Simpson AJ. Biparental mitochondrial DNA inheritance in the parasitic trematode Schistosoma mansoni. *The Journal of Parasitology*, 87:79–82, 2001.

Jarne P and David P. Microsatellites, Transposable Elements and the X Chromosome. *Molecular Biology and Evolution*, 15:28–34, 1998.

Jenuth JP, Peterson A, Fu K, and Shoubridge E. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature Genetics*, pages 146–151, 1996.

Jeyaprakash A and Hoy MA. The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. *Gene*, 391:264–274, 2007.

Jiggins FM. Male-killing Wolbachia and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, 164:5–12, 2003.

Jungblut B and Sommer RJ. Novel cell-cell interactions during vulva development in *Pristionchus pacificus*. *Development*, 127:3295–3303, 2000.

Jurka J and Pethiyagoda C. Simple repetitive DNA sequences from primates: compilation and analysis. *Journal of Molecular Evolution*, 40:120–126, 1995.

Kaneda H, Hayashi J, Takahama S, Taya C, Lindahl KF, and Yonekawa H. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proceedings of the National Academy of Sciences*, 92:4542–4546, 1995.

Kang S, Sultana T, Eom KS, Park Yung Chul, Soonthornpong Nathan, Nadler SA, and Park J-K. The

mitochondrial genome sequence of *Enterobius vermicularis* (Nematoda: Oxyurida)–an idiosyncratic gene order and phylogenetic information for chromadorean nematodes. *Gene*, 429:87–97, 2009.

Katti MV, Ranjekar PK, and Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution*, 18:1161–1167, 2001.

Kayser M, Vowles EJ, Kappei D, and Amos W. Microsatellite length differences between humans and chimpanzees at autosomal Loci are not found at equivalent haploid Y chromosomal Loci. *Genetics*, 173: 2179–2186, 2006.

Keddie EM, Higazi T, and Unnasch TR. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Molecular and Biochemical Parasitology*, 95:111–27, 1998.

Kenning C, Kipping I, and Sommer RJ. Isolation of mutations with dumpy-like phenotypes and of collagen genes in the nematode *Pristionchus pacificus*. *Genesis*, 40:176–83, 2004.

Kikuchi T, Aikawa T, Kosaka H, Pritchard L, Ogura N, and Jones JT. Expressed sequence tag (EST) analysis of the pine wood nematode Bursaphelenchus xylophilus and B. mucronatus. *Molecular and Biochemical Parasitology*, 155:9–17, 2005.

Kim Kyu-Heon, Eom Keeseon S, and Park Joong-Ki. The complete mitochondrial genome of *Anisakis simplex* (Ascaridida: Nematoda) and phylogenetic implications. *International journal for parasitology*, 36: 319–28, 2006.

Kimura M. Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution*, 1:84–93, 1983.

Kimura M and Crow JF. The Number of Alleles that Can Be Maintained in a Finite Population. *Genetics*, 49: 725–738, 1964.

Kiontke K, Barrière A, Kolotuev I, Podbilewicz B, Sommer RJ, Fitch DHA, and Félix M-A. Trends, stasis, and drift in the evolution of nematode vulva development. *Current Biology*, 17:1925–1937, 2007.

Kitazaki K and Kubo T. Cost of Having the Largest Mitochondrial Genome: Evolutionary Mechanism of Plant Mitochondrial Genome. *Journal of Botany*, 2010:1–12, 2010.

Koehler CM, Lindberg GL, Brown DR, Beitz DC, Freeman AE, Mayfield JE, and Myers AM. Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics*, 129:247–255, 1991.

Kofler R, Schlötterer C, Luschützky E, and Lelley T. Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. *BMC Genomics*, 9:612, 2008.

Kolpakov R. mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research*, 31: 3672–3678, 2003.

Kruglyak S, Durrett RT, Schug MD, and Aquadro CF. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences*, 95:10774–10778, 1998.

Kruglyak S, Durrett R, Schug MD, and Aquadro CF. Distribution and Abundance of Microsatellites in the Yeast Genome Can Be Explained by a Balance Between Slippage Events and Point Mutations. *Molecular Biology and Evolution*, 17:1210–9, 2000.

Kumar S and Subramanian S. Mutation rates in mammalian genomes. *Proceedings of the National Academy*

of Sciences, 99:803–808, 2002.

Kvist L. Paternal Leakage of Mitochondrial DNA in the Great Tit (*Parus major*). *Molecular Biology and Evolution*, 20:243–247, 2003.

Ladoukakis ED and Zouros E. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution*, 18:1168–1175, 2001.

Lai Y and Sun F. The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. *Molecular Biology and Evolution*, 20:2123–2131, 2003.

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, and Gray MW. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, 387:493–497, 1997.

Lang BF, Gray MW, and Burger G. Mitochondrial Genome Evolution and the Origin of Eukaryotes. *Annual Review of Genetics*, 33:351–397, 1999.

Lavrov DV and Brown WM. *Trichinella spiralis* mtDNA: A Nematode Mitochondrial Genome That Encodes a Putative ATP8 and Normally Structured tRNAs and Has a Gene Arrangement Relatable to Those of Coelomate Metazoans. *Genetics*, 157:621–637, 2001.

Leclercq S, Rivals E, and Jarne P. Detecting microsatellites within genomes: significant variation among algorithms. *BMC Bioinformatics*, 8:125, 2007.

Levinson G and Gutman GA. Slipped-Strand Mispairing: A Major Mechanism for DNA Sequence Evolution. *Molecular Biology and Evolution*, 4:203–221, 1987.

Li M-W, Lin R-Q, Song H-Q, Wu X-Y, and Zhu X-Q. The complete mitochondrial genomes for three *Toxocara* species of human and animal health significance. *BMC genomics*, 9:224, 2008.

Li Y-C, Korol AB, Fahima T, Beiles A, and Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11:2453–2465, 2002.

Litt M and Luty JA. A Hypervariable Microsatellite Revealed by *in vitro* Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. *American Journal of Human Genetics*, 44:397–401, 1989.

Lohe AR, Hilliker AJ, and Roberts PA. Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics*, 134:1149–1174, 1993.

Lunt DH and Hyman BC. Animal mitochondrial DNA recombination. *Nature*, 387:247, 1997.

Lynch M. Mutation Accumulation in Nuclear, Organelle, and Prokaryotic Transfer RNA genes. *Molecular Biology and Evolution*, 14:914–925, 1997.

Lynch M. The origins of eukaryotic gene structure. *Molecular Biology and Evolution*, 23:450–68, 2006a.

Lynch M. Streamlining and simplification of microbial genome architecture. *Annual Review of Microbiology*, 60:327–349, 2006b.

Lynch M. *The Origins of Genome Architecture*. Sinauer Associates, 2007.

Lynch M and Blanchard JL. Deleterious mutation accumulation in organelle genomes. *Genetica*, 102-103: 29–39, 1998.

Lynch M and Milligan BG. Analysis of population genetic structure with RAPD markers. *Molecular Ecology*,

3:91–99, 1994.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson W, Okamoto K, Kulkarni S, Hartl DL, and Thomas W. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, 105:9272–9277, 2008.

MacAlpine DM, Kolesar J, Okamoto K, Butow RA, and Perlman PS. Replication and preferential inheritance of hypersupressive petite mitochondrial DNA. *The EMBO Journal*, 20:1807–1817, 2001.

Macaulay V, Richards M, and Sykes B. Mitochondrial DNA recombination-no need to panic. *Proceedings of the Royal Society Biology Section*, 266:2037–2039, 1999.

Marchington DR, Macaulay V, Hartshorne GM, Barlow D, and Poulton J. Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. *American Journal of Human Genetics*, 63:769–775, 1998.

Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, and Kelly JK. Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity*, 103:310–317, 2009.

Martin AP, Naylor GJ, and Palumbi SR. Rates of Mitochondrial DNA Evolution in Sharks are Slow Compared with Mammals. *Nature*, 357:153–155, 1992.

Mayer MG and Sommer RJ. Natural Variation in *Pristionchus pacificus* Dauer Formation Reveals Cross-Preference Rather than Self-Preference of Nematode Dauer Pheromones. *Proceedings of the Royal Society Biology Section*, 278:2784–2790, 2011.

Mayer WE, Herrmann M, and Sommer RJ. Phylogeny of the nematode genus *Pristionchus* and implications for biodiversity, biogeography and the evolution of hermaphroditism. *BMC Evolutionary Biology*, 7:104, 2007.

Mayer WE, Herrmann M, and Sommer RJ. Molecular phylogeny of beetle associated diplogastrid nematodes suggests host switching rather than nematode-beetle coevolution. *BMC Evolutionary Biology*, 9:212, 2009.

Mayer WE, Schuster LN, Bartelmes G, Dieterich C, and Sommer RJ. Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evolutionary Biology*, 11:13, 2011.

Maynard Smith J and Szathmáry E. *The Major Transitions in Evolution.* Oxford University Press, 1995.

Messier W, Li SH, and Stewart CB. The birth of microsatellites. *Nature*, 381:483, 1996.

Meusel MS and Moritz RF. Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. *Current Genetics*, 24:539–543, 1993.

Milbury CA and Gaffney PM. Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. *Marine Biotechnology*, 7:697–712, 2005.

Molnar RI, Bartelmes G, Dinkelacker I, Witte H, and Sommer RJ. Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*. *Molecular Biology and Evolution*, 28: 2317–2326, 2011.

Montiel R, Lucena MA, Medeiros J, and Simões N. The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae*: insights into nematode mitochondrial DNA evolution and

phylogeny. *Journal of Molecular Evolution*, 62:211–25, 2006.

Morgan K, McGaughran A, Villate L, Herrmann M, Witte H, Bartelmes G, Rochat J, and Sommer RJ. Multi locus analysis of *Pristionchus pacificus* on La Réunion Island reveals an evolutionary history shaped by multiple introductions, constrained dispersal events and rare out-crossing. *Molecular Ecology*, 21: 250–266, 2012.

Moritz C, Dowling TE, and Brown WM. Evolution of Animal Mitochondrial DNA: Relevance for Population Biology and Systematics. *Annual Review of Ecology and Systematics*, 18:269–292, 1987.

Morris JC, Drew ME, Klingbeil MM, Motyka SA, Saxowsky TT, Wang Z, and Englund PT. Replication of kinetoplast DNA: an update for the new millennium. *International Journal for Parasitology*, 31:453–458, 2001.

Myers TR and Greenwald I. Wnt signal from multiple tissues and lin-3 / EGF signal from the gonad maintain vulval precursor cell competence in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 104:20368–20373, 2007.

Nabholz B, Glémin S, and Galtier N. Strong variations of mitochondrial mutation rate across mammals–the longevity hypothesis. *Molecular Biology and Evolution*, 25:120–130, 2008.

Nabholz B, Glémin S, and Galtier N. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology*, 9:54, 2009.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Houm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, and White R. Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping. *Science*, 235:1616–1622, 1987.

Noor MAF, Kliman RM, and Machado CA. Evolutionary History of Microsatellites in the Obscura Group of *Drosophila*. *Molecular Biology and Evolution*, 18:551–556, 2001.

Normark BB. Molecular systematics and evolution of the aphid family Lachnidae. *Molecular Phylogenetics and Evolution*, 14:131–140, 2000.

Ogawa A, Streit A, Antebi A, and Sommer RJ. A Conserved Endocrine Mechanism Controls the Formation of Dauer and Infective Larvae in Nematodes. *Current Biology*, 19:67–71, 2009.

Ogawa A, Bento G, Bartelmes G, Dieterich C, and Sommer RJ. *Pristionchus pacificus daf-16* is essential for dauer formation but dispensable for mouth form dimorphism. *Development*, 138:1281–4, 2011.

Ohta T and Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical research*, 22:201–4, 1973.

Ojala D, Montaya J, and Attardi G. tRNA punctuation model of RNA porcessing in human mitochondria. *Nature*, 290, 1981.

Okimoto R, Macfarlane JL, and Wolstenholme DR. Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. *Nucleic Acids Research*, 18:6113–6118, 1990.

Okimoto R, Chamberlin HM, Macfarlane JL, and Wolstenholme DR. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and

potential for host-race identification. *Nucleic Acids Research*, 19:1619–1626, 1991.

Okimoto R, Macfarlane JL, Clary D, and Wolstenholme DR. The Mitochondrial Genomes of Two Nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics*, 130:471–498, 1992.

Oldenburg DJ and Bendich AJ. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *Journal of Molecular Biology*, 335:953–970, 2004.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, and Vieira LMC. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29:294–307, 2006.

Papetti C, Liò P, Rüber L, Patarnello T, and Zardoya R. Antarctic fish mitochondrial genomes lack ND6 gene. *Journal of Molecular Evolution*, 65:519–528, 2007.

Pérez M, Cruz F, and Presa P. Distribution properties of polymononucleotide repeats in molluscan genomes. *The Journal of Heredity*, 96:40–51, 2005.

Piganeau G and Eyre-Walker A. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity*, 92:282–288, 2004.

Pires-daSilva A and Sommer RJ. Conservation of the global sex determination gene *tra-1* in distantly related nematodes. *Genes and Development*, 18:1198–1208, 2004.

Posada D and Crandall KA. The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of Molecular Evolution*, 54:396–402, 2002.

Poulton J, Macaulay V, and Marchington DR. Mitochondrial genetics '98 is the bottleneck cracked? *American Journal of Human Genetics*, 62:752–757, 1998.

Poyton RO and McEwen JE. Crosstalk between nuclear and mitochondrial genomes. *Annual Review of Biochemistry*, 65:563–607, 1996.

Primmer CR and Ellegren H. Patterns of Molecular Evolution in Avian Microsatellites. *Molecular Biology and Evolution*, 15:997–1008, 1998.

Primmer CR, Saino N, Mø ller AP, and Ellegren H. Directional evolution in germline microsatellite mutations. *Nature Genetics*, 13:391–393, 1996.

Pupko T and Graur D. Evolution of Microsatellites in the Yeast *Saccharomyces cerevisiae*: Role of Length and Number of Repeated Units. *Journal of Molecular Evolution*, 48:313–316, 1999.

Rand D. Mitochondrial Genomics Flies High. *Trends in Ecology and Evolution*, 16:2–4, 2001.

Rand DM. Mitigating mutational meltdown in mammalian mitochondria. *PLoS Biology*, 6:e35, 2008.

Richard G-F and Pâques F. Mini- and microsatellite expansions: the recombination connection. *EMBO Reports*, 1:122–126, 2000.

Riddle DL and Albert P. Genetic and environmental regulation of dauer larva development. In Riddle DL, Blumenthal T, Meyer BJ, and Preiss JR, editors, *C. elegans II*, pages 739–768. Cold Spring Harbor Laboratory Press, New York, 1997.

Robertson BC, Steeves TE, McBride KP, Goldstien SJ, Williams M, and Gemmell NJ. Phylogeography of the New Zealand blue duck (*Hymenolaimus malacorhynchos*): implications for translocation and species recovery. *Conservation Genetics*, 8:1431–1440, 2007.

Rödelsperger C and Sommer RJ. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. *BMC Evolutionary Biology*, 11:239, 2011.

Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, and Garcia-Mas J. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics*, 12:424, 2011.

Roest Crollius H, Jaillon O, Dasilva C, Ozouf-Costaz C, Fizames C, Fischer C, Bouneau L, Billault A, Quetier F, Saurin W, Bernot A, and Weissenbach J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome research*, 10:939–49, 2000.

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

Rolfsmeier ML and Lahue RS. Stabilizing effects of interruptions on trinucleotide repeat expansions in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 20:173–180, 2000.

Rose O and Falush D. A Threshold Size for Microsatellite Expansion. *Molecular Biology and Evolution*, 15: 613–615, 1998.

Rosso MN, Favery B, Piotte C, Arthaud L, de Boer JM, Hussey RS, Bakker J, Baum TJ, and Abad P. Isolation of a cDNA encoding a beta-1,4-endoglucanase in the rootknot nematode Meloidogyne incognita and expression analysis during plant parasitism. *Molecular Plant-Microbe Interactions*, 12:585–591, 1999.

Rudel D, Tian H, and Sommer RJ. Wnt signaling in *Pristionchus pacificus* gonadal arm extension and the evolution of organ shape. *Proceedings of the National Academy of Sciences*, 105:10826–10831, 2008.

Sagan L. On the origin of mitosing cells. *Journal of Theoretical Biology*, 14:225–274, 1967.

Sainudiin R, Durrett RT, Aquadro CF, and Nielsen R. Microsatellite mutation models: insights from a comparison of humans and chimpanzees. *Genetics*, 168:383–395, 2004.

Sammler S, Bleidorn C, and Tiedemann R. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics*, 12:35, 2011.

Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, and Becker PB. Dual regulation of the *Drosophila hsp26* promoter *in vitro*. *Nucleic Acids Research*, 23:2479–2487, 1995.

Saraste M. Oxidative phosphorylation at the fin de siècle. *Science*, 283:1488–1493, 1999.

Satoh M and Kuroiwa T. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research*, 196:137–140, 1991.

Schierup MH and Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics*, 156:879–891, 2000.

Schlager B, Roseler W, Zheng M, Gutierrez A, and Sommer RJ. HAIRY-like transcription factors and the evolution of the nematode vulva equivalence group. *Current Biology*, 16:1386–1394, 2006.

Schlager B, Wang X, Braach G, and Sommer RJ. Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis*, 47: 300–304, 2009.

Schlötterer C. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109:365–371, 2000.

Schlötterer C, Ritter R, Harr B, and Brem G. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution*, 15: 1269–74, 1998.

Schug MD, Mackay TFC, and Aquadro CF. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genetics*, pages 99–102, 1997.

Schwartz M and Vissing J. Paternal inheritance of mitochondrial DNA. *The New England Journal of Medicine*, 347:576–580, 2002.

Seyfert AL, Cristescu MEA, Frisse L, Schaack S, Thomas WK, and Lynch M. The Rate and Spectrum of Microsatellite Mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics*, 178:2113–2121, 2008.

Shao R and Barker SC. Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. *Parasitology*, 134:153–167, 2007.

Shao R, Mitani H, Barker SC, Takahashi M, and Fukunaga M. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *Journal of Molecular Evolution*, 60:764–773, 2005.

Sharma PC, Grover A, and Kahl G. Mining microsatellites in eukaryotic genomes. *Trends in Biotechnology*, 25:490–498, 2007.

Shinde D. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Research*, 31:974–980, 2003.

Sinha A, Sommer RJ, and Dieterich C. Divergent gene expression in the conserved dauer stage of the nematode *Pristionchus pacificus* and *Caenorhabditis elegans*. *BMC Genomics*, submitted, 2012.

Smant G, Stokkermans J, Yan YT, de Boer JM, Baum TJ, Wang XH, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, and Bakker J. Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences*, 95:4906–4911, 1998.

Smith DR and Lee RW. Mitochondrial Genome of the Colorless Green Alga *Polytomella capuana*: A Linear Molecule with an Unprecedented GC Content. *Molecular Biology and Evolution*, 25:487–496, 2007.

Sommer RJ. Evolution of nematode development. *Current Opinion in Genetics and Development*, 10:443–448, 2000.

Sommer RJ. As good as they get: cells in nematode vulva development. *Current Opinion in Cell Biology*, 13: 715–720, 2001.

Sommer RJ. The future of evo-devo model systems and evolutionary theory. *Nature Reviews Genetics*, 10: 416–421, 2009.

Sommer RJ and Sternberg PW. Evolution of Nematode Vulval Fate Patterning. *Developmental Biology*, 173: 396–407, 1996.

Sommer RJ, Carta LK, Kim S-Y, and Sternberg PW. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda : Neodiplogastridae). *Fundamental and Applied Nematology*, 19: 511–521, 1996.

Sprinzl M, Hartmann T, Weber J, Blank J, and Zeidler R. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 26:148–153, 1989.

Stearns SC and Hoekstra RF. *Evolution: an introduction.* Oxford University Press, New York, 2nd edition, 2005.

Sudhaus W. Preadaptive plateau in Rhabditida (Nematoda) allowed the repeated evolution of zooparasites, with an outlook on evolution of life cycles within Spiroascarida. *Paleodiversity*, Supplement:117–130, 2010.

Sudhaus W and Fürst von Lieven A. A phylogenetic classification and catalogue of the Diplogastridae (Secernentea, Nematoda). *Journal of Nematode Morphology and Systematics*, 6:43–89, 2003.

Sulston JE and Brenner S. The DNA of *Caenorhabditis elegans*. *Genetics*, 77:95–104, 1974.

Sureshkumar S, Todesco M, Schneeberger K, Harilal R, Balasubramanian S, and Weigel D. A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*. *Science*, 323:1060–1063, 2009.

Sutovsky P, Van Leyen K, McCauley T, Day BN, and Sutovsky M. Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. *Reproductive BioMedicine Online*, 8:24–33, 2004.

Szathmáry E and Maynard Smith J. The major evolutionary transitions. *Nature*, 374:227–232, 1995.

Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta*, 1410:103–23, 1999.

Tachida H and Iizuka M. Persistence of Repeated Sequences That Evolve by Replication Sippage. *Genetics*, 131:471–478, 1992.

Tang S and Hyman BC. Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. *Genetics*, 176:1139–1150, 2007.

Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17:6463–6471, 1989.

Tautz D, Trick M, and Dover GA. Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, 322:652–656, 1986.

Taylor JS, Durkin JMH, and Breden F. The Death of a Microsatellite: A Phylogenetic Perspective on Microsatellite Interruptions. *Molecular Biology and Evolution*, 16:567–572, 1999.

Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, and Sing CF. Recombinational and mutational hotspots within the human lipoprotein lipase gene. *American Journal of Human Genetics*, 66: 69–83, 2000.

The *C. elegans* Sequencing Consortium . Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science*, 282:2012–2018, 1998.

Thiel T, Michalek W, Varshney RK, and Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 106:411–422, 2003.

Thyagarajan B, Padua RA, and Campbell C. Mammalian mitochondria possess homologous DNA recom-

bination activity. *The Journal of Biological Chemistry*, 271:27536–27543, 1996.

Tian H, Schlager B, Xiao H, and Sommer RJ. Wnt signaling induces vulva development in the nematode *Pristionchus pacificus*. *Current Biology*, 18:142–146, 2008.

Tijsterman M, Pothof J, and Plasterk RHA. Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient *Caenorhabditis elegans*. *Genetics*, 161:651–660, 2002.

Toth G, Gaspari Z, and Jurka J. Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Research*, 10:967–981, 2000.

Unseld M, Marienfeld JR, Brandt P, and Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, 15:57–61, 1997.

Valdes AM, Slatkin M, and Freimer NB. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, 133:737–749, 1993.

Van Der Veer M and De Vries E. A single nucleotide polymorphism map of the mitochondrial genome of the parasitic nematode *Cooperia oncophora*. *Parasitology*, 128:421–431, 2004.

van Megen H, van den Elsen S, Holterman M, Karssen G, Mooyman P, Bongers T, Holovachov O, Bakker J, and Helder J. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology*, 11:927–950, 2009.

van Oppen MJH, Catmull J, McDonald BJ, Hislop NR, Hagerman J, and Miller DJ. The mitochondrial genome of *Acropora tenuis* (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. *Journal of Molecular Evolution*, 55:1–13, 2002.

Varela MA and Amos W. Heterogeneous distribution of SNPs in the human genome: microsatellites as predictors of nucleotide diversity and divergence. *Genomics*, 95:151–159, 2010.

Vassilieva LL and Lynch M. The Rate of Spontaneous Mutation for Life-History Traits in *Caenorhabditis elegans*. *Genetics*, 151:119–29, 1999.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, and Doebley J. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, 19:1251–1260, 2002.

Wai T, Teoli D, and Shoubridge EA. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nature Genetics*, 40:1484–1488, 2008.

Wang M and Sternberg PW. Pattern formation during *C. elegans* vulval induction. *Current Topics in Developmental Biology*, 51:189–220, 2001.

Wang X and Sommer RJ. Antagonism of LIN-17/Frizzled and LIN-18/Ryk in Nematode Vulva Induction Reveals Evolutionary Alterations in Core Developmental Pathways. *PLoS Biology*, 9:e1001110, 2011.

Watanabe K and Yokobori S-I. tRNA Modification and Genetic Code Variations in Animal Mitochondria. *Journal of Nucleic Acids*, 2011, 2011.

Weinreich DM and Rand DM. Contrasting Patterns of Nonneutral Evolution in Proteins Encoded in Nuclear and Mitochondrial Genomes. *Genetics*, 156:385–399, 2000.

Weller AM, Mayer WE, Rae R, and Sommer RJ. Quantitative assessment of the nematode fauna present on Geotrupes dung beetles reveals species-rich communities with a heterogeneous distribution. *The*

Journal of Parasitology, 96:525–531, 2010.

White DJ, Wolff JN, Pierson M, and Gemmell NJ. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17:4925–4942, 2008.

Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, and Sibly RM. Likelihood-based estimation of microsatellite mutation rates. *Genetics*, 164:781–787, 2003.

Wiesner RJ, Rüegg JC, and Morano I. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochemical and Biophysical Research Communications*, 183:553–559, 1992.

Wilder J and Hollocher H. Mobile Elements and the Genesis of Microsatellites in Dipterans. *Molecular Biology and Evolution*, 18:384–392, 2001.

Wolff JN and Gemmell NJ. Estimating mitochondrial DNA content of chinook salmon spermatozoa using quantitative real-time polymerase chain reaction. *Biology of Reproduction*, 79:2472–2452, 2008.

Wolstenholme DR. Genetic novelties in mitochondrial genomes of multicellular animals. *Current Opinion in Genetics and Development*, 2:918–925, 1992.

Wolstenholme DR and Jeon KW, editors. *International Review of Cytology - A Survey of Cell Biology - Mitochondrial Genomes*. Academic Press, 1992. ISBN 0123645441.

Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, and Wahleithner JA. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proceedings of the National Academy of Sciences*, 84:1324–8, 1987.

Xu S, Schaack S, Seyfert A, Choi E, Lynch M, and Cristescu ME. High mutation rates in the mitochondrial genomes of *Daphnia pulex*. *Molecular Biology and Evolution*, page advanced cess, 2011.

Yan YT, Smant G, Stokkermans J, Qin L, Helder J, Baum TJ, Schots A, and Davis E. Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. *Gene*, 220:61–70, 1998.

Zauner H and Sommer RJ. Evolution of robustness in the signaling network of *Pristionchus* vulva development. *Proceedings of the National Academy of Sciences*, 104:10086–100891, 2007.

Zauner H, Mayer WE, Herrmann M, Weller AM, Erwig M, and Sommer RJ. Distinct patterns of genetic variation in *Pristionchus pacificus* and *Caenorhabditis elegans*, two partially selfing nematodes with cosmopolitan distribution. *Molecular Ecology*, 16:1267–1280, 2007.

Zheng M, Messerschmidt D, Jungblut B, and Sommer RJ. Conservation and diversification of Wnt signaling function during the evolution of nematode vulva development. *Nature Genetics*, 37:300–304, 2005.

Zhuang X and Cheng C-HC. ND6 gene "lost" and found: evolution of mitochondrial gene rearrangement in Antarctic notothenioids. *Molecular Biology and Evolution*, 27:1391–1403, 2010.

Zouros E, Freeman KR, Ball AO, and Pogson GH. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature*, 359:412–414, 1992.

Zouros E, Oberhauser Ball A, Saavedra C, and Freeman KR. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences*, 91:7463–7467, 1994.

Articles

Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*

Ruxandra I. Molnar,[1] Gabi Bartelmes,[1] Iris Dinkelacker,[1] Hanh Witte,[1] and Ralf J. Sommer*[,1]

[1]Department for Evolutionary Biology, Max-Planck Institute for Developmental Biology, Tübingen, Germany

***Corresponding author:** E-mail: ralf.sommer@tuebingen.mpg.de.

Associate editor: Patricia Beldade

Abstract

Evolutionary reconstruction of the natural history of an organism ultimately requires knowledge about the development, population genetics, ecology, and phylogeny of the species. Such investigations would benefit from studies of mutational processes because mutations are the source of natural variation. The nematode *Pristionchus pacificus* has been developed as a model organism in evolutionary biology by comparing its development with *Caenorhabditis elegans*. *Pristionchus pacificus* and related species are associated with scarab beetles, and their ecology and phylogeny are well known. More than 200 *P. pacificus* isolates from all over the world are available for this cosmopolitan species. We generated mutation accumulation (MA) lines in *P. pacificus* to study spontaneous mutation rates in the mitochondrial genome and compared mutation rate estimates with natural variation between nine representative isolates of the species. The *P. pacificus* mitochondrial genome is 15,955 bp in length and is typical for nematodes. *Pristionchus pacificus* has all known mitochondrial genes and contains an unusual suppressor transfer RNA (tRNA) for the codon UAA. This has most likely influenced the spectrum of observable mutations because 6 of 12 mutations found in the 82 MA lines analyzed are nonsense mutations that can be suppressed by the suppressor tRNA. The overall mutation rate in *P. pacificus* is 7.6×10^{-8} per site per generation and is less than one order of magnitude different from estimates in *C. elegans* and *Drosophila*. Using this mutation rate estimate in a comparison of the mitochondrial genome of nine *P. pacificus* isolates, we calculate the minimum time to the most recent common ancestor at 10^5–10^6 generations. The combination of mutation rate analysis with intraspecific divergence provides a powerful tool for the reconstruction of the natural history of *P. pacificus*, and we discuss the ecological implication of these findings.

Key words: mutation accumulation lines, *Pristionchus pacificus*, mitochondrial DNA, divergence times, *Caenorhabditis elegans*.

Introduction

The appearance of new mutations and the analysis of mutation rates are crucial for any attempt to build a comprehensive understanding of the evolution of animal and plant species (Lynch 2007). In combination with phylogenetics, population genetics, and ecology, knowledge of mutational processes can provide the basis for estimating the rate of molecular evolution, the effective population size, and the natural history of a given species (Charlesworth 2009). Indirect estimates of mutational rates depend on levels of divergence or polymorphisms and assume that mutations are effectively neutral or need the quantification of diverse fitness traits in mutation accumulation (MA) line experiments (see e.g., Vassilieva and Lynch 1999; Baer et al. 2005). With MA line experiments, it is also possible to make direct estimates of minimum mutation rates (Lynch 2010). These experiments are limited to a small number of experimental model organisms, such as the rhabditid nematodes *Caenorhabditis elegans* (Denver et al. 2000, 2004) and *C. briggsae* (Baer et al. 2005; Howe et al. 2010), *Drosophila melanogaster* (Mukai 1964; Haag-Liautard et al. 2007, 2008), *Daphnia pulex* (Seyfert et al. 2008), *Arabidopsis thaliana* (Ossowski et al. 2010), *Escherichia coli* (Barrick et al. 2009), and *Saccharomyces cerevisiae* (Lynch et al. 2008). In a MA line experimental setup, the population suffers a drastic bottleneck for many generations, assuring that all but the most deleterious mutations accumulate in a nearly neutral fashion (Vassilieva and Lynch 1999). Herein, we have used a MA line experiment in the nematode *Pristionchus pacificus* and study mutations in the mitochondrial genome. We use the frequency of mitochondrial mutations obtained after 142 generations of inbreeding in 82 MA lines to interpret the natural variation of the corresponding mitochondrial sequences in nine *P. pacificus* wild isolates.

Animal mitochondrial genomes differ from nuclear DNA in many ways, with hypermutability being the most striking character (Lynch 2007). The mitochondrial DNA (mtDNA) evolves rapidly in population, and it is usually transmitted maternally without intermolecular recombination (but see Zouros et al. 1992; Howe et al. 2010), and therefore, it has been intensively used for phylogeographic studies (Avise 1998) for more than three decades. Usually, mitochondrial mutation rates are one order of magnitude higher than mutation rates in the nuclear genome. These high mutation rates resulted in the development of mtDNA as predominant marker for studies of biodiversity, phylogeny, and natural variation between populations. Metazoan mtDNA is usually a circular molecule between 14 and

Table 1. Mitochondrial Genome Size and Gene Content Overview of Representative Nematodes.

Nematode	Clade: Blaxter et al. (1998)	Clade: Megen et al. (2009)	Size (bp)	Noncoding Nucleotides (bp)[a]	Gene Content: PCGs	Gene Content: tRNAs	Gene Content: rRNAs	Reference
Pristionchus pacificus	V	9A	15,955	2,772	12	22 + $tRNA^{Sup}$	2	Present study
Caenorhabditis elegans	V	9A	13,794	575	12	22	2	Okimoto et al. (1992)
Ancylostoma duodenale	V	9B	13,721	513	12	22	2	Hu et al. (2002)
Necator americanus	V	9B	13,604	387	12	22	2	Hu et al. (2002)
Steinernema carpocapsae	IV	10A	13,925	696	12	22	2	Montiel et al. (2006)
Strongyloides stercoralis	IV	10B	13,758	575	12	22	2	Hu et al. (2003)
Meloidogyne javanica	IV	12B	20,565	7,000	12	22	2	Okimoto et al. (1991)
Radopholus similis[b]	IV	12B	16,791	3,904	12	22	2	Jacob et al. (2009)
Onchocerca volvulus	III	8B	13,747	357	12	22	2	Keddie et al. (1998)
Ascaris suum	III	8B	14,284	1,048	12	22	2	Okimoto et al. (1992)
Trichinella spiralis[c]	I	2A	21,000–24,000	8,098–10,098	13	22 + $tRNA^{Met2}$	2	Lavrov and Brown (2001)
Xiphinema americanum[c]	I	2C	12,626	305	12	17	2	He et al. (2005)

NOTE.—PCG, protein-coding genes.
[a] Sum of the bp included in the D-loop and other noncoding nucleotide regions longer than 10 bp.
[b] $UAA\text{-}tRNA^{STOP}$ to $UAA\text{-}tRNA^{Tyr}$ reassignment
[c] The mtDNA is transcribed bidirectional.

42 kb containing 36–37 genes, including 12–13 protein-coding genes for the various subunits of the enzyme complexes of the respiratory chain, two ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes (Wolstenholme 1992). Despite this generally conserved organization, there is substantial variation in size, which is attributable to differences in the noncoding regions. The mtDNA contains at least one large noncoding region called the D-loop, or control region, which contains the origin of replication of the DNA molecule. Nematode mitochondrial genomes are very compact and contain relatively short rRNA molecules and truncated tRNAs (table 1). The 45 complete nematode mtDNAs sequenced to date show a large proportion of variability: The gene *atp8* is absent from all the mtDNAs of sampled nematodes, except one species, *Trichinella spiralis* (Lavrov and Brown 2001); *Globodera pallida* presents a multipartite mitochondrial genome (Armstrong et al. 2000); pseudogenes are present in *C. briggsae* (Howe and Denver 2008); and repeated genes are present in the mitochondrial genome of *Romanomermis culicivorax*, which also exhibits an exceptionally large mtDNA of 26–32 kb (Hyman and Azevedo 1996).

The diplogastrid nematode *P. pacificus* has been developed as a model system in evolutionary biology based on its short generation time, hermaphroditic, predominantly selfing mode of reproduction, easy laboratory culture, and other technical advantages, which allow a detailed comparison with *C. elegans* (Sommer et al. 1996; Hong and Sommer 2006). With the availability of forward and reverse genetics and DNA-mediated transformation (Schlager et al. 2009), functional studies of many biological processes are feasible. Original work concentrated on evolutionary developmental biology (evo-devo), more specifically on the signaling pathways involved in the development of the vulva (Zheng et al. 2005; Tian et al. 2008). More recent studies on *P. pacificus* extend to other areas in evolutionary biology with the aim to link evo-devo with population genetics and ecology (Sommer 2009). Phylogenetic studies have robustly inferred the position of *P. pacificus* within the family Diplogastridae (Mayer et al. 2009) and within the genus *Pristionchus* (Mayer et al. 2007).

Ecological studies in *Pristionchus* have shown that these nematodes live in close association with scarab beetles (Herrmann et al. 2006a). Strains of *P. pacificus* have been found on *Exomala orientalis* (Oriental beetle) in Japan and the United States (Herrmann et al. 2007) and, more recently, on *Oryctes borbonicus* (Rhinoceros beetle) and other scarab beetles on La Réunion in the Indian Ocean (Herrmann et al. 2010). A collection of more than 200 *P. pacificus* isolates from around the world is currently available for laboratory study. To provide a better account of the natural history of *P. pacificus*, we compared the mtDNA of nine *P. pacificus* strains and combined this knowledge of mitochondrial genetic variation with MA line–derived mutation rate estimates. We evaluate the mutation rate in the MA lines and derive dates of divergence by applying the internal molecular clock calibration to substitution rates inferred for the mitochondrial genome of the *P. pacificus* strains.

Materials and Methods

Strains

A list of all *P. pacificus* strains used in this study is provided in table 2.

MA Lines

MA lines were generated as depicted in figure 1, following the protocol of Vassilieva and Lynch (1999), from a single individual of the wild-type *P. pacificus* strain from California (PS312). Worms were cultured at 20 °C on nematode growth medium seeded with *E. coli* as a food source, as originally described for *P. pacificus* (Sommer et al. 1996). This line has been maintained in the lab for many generations

Table 2. Strains of *Pristionchus pacificus* Used in This Study.

	Sample Location	Year	Sample Type	Reference
PS312	Pasadena (California), USA	1988	Soil	Sommer et al. (1996)
RS5416	Saint Benoit, La Réunion	2009	*Maladera affinis*	Herrmann et al. (2010)
RS5211	Japan	2006	*Exomala orientalis*	Herrmann et al. (2010)
RS5282	Antalya, Turkey	2007	Soil	Herrmann et al. (2010)
JU150	Antananarivo, Madagascar	1999	Soil	Zauner and Sommer (2007)
RS5399	Trois Bassins, La Réunion	2009	*Oryctes borbonicus*	Herrmann et al. (2010)
RS5361	Nez du Boef, La Réunion	2010	Soil	Herrmann et al. (2010)
RS5275	Santa Cruz, Bolivia	2006	Scarab beetles	Hong et al. (2008)
RS5340	Basse Vallé, La Réunion	2008	*Adoretus* sp.	Herrmann et al. (2010)

before the experiment. We initiated the experiment by establishing 100 clonal lines from a single hermaphrodite that was inbred for three generations. Each MA line was propagated across 142 generations as a single randomly picked L3 stage worm from the middle of the reproductive period of the previous generation. Two previous generations for each line were kept at 15 °C in order to prevent accidental loss of the lines. From the original 100 lines, 82 survived until the 142nd generation. All the lines were frozen at generations 1, 27, 70, and 142.

DNA Extraction, Amplification, and Sequencing

For each MA line, we prepared genomic DNA from two plates of worms using worm lysis buffer (50 mM KCl, 10 mM Tris–HCl pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, and 5 μg/ml proteinase K) and incubated the suspension for 2 h at 65 °C, followed by inactivation of the proteinase K at 95 °C for 10 min. The DNA was diluted to approximate 25 ng/μl for the polymerase chain reactions (PCRs). The entire mtDNA (except the D-loop) was amplified from total genomic DNA following two procedures: 1) small amplicons (500–1,000 bp) in 20 μl reactions using Taq (New England BioLabs) and 2) long amplicons (2,500–3,000 bp) using Expand Long Range dNTPack (Roche). A list of primers is available in supplementary table S1, Supplementary Material online. If necessary, additional primers were used to ensure reliable amplification of the entire mtDNA. The small amplicons were sequenced with the PCR primers using a ABI3730*xl* capillary platform. We used the software SeqMan from DNAStar Inc. to visualize the trace files, align the trace files, and assemble the mtDNA for each line independently. The long PCR products were visualized on 1.5% agarose gels to ensure amplification success and screen for large heteroplasmic deletion events. For the *P. pacificus* strains, we prepared DNA from ten worms in 20 μl of lysis buffer with 160 μl final volume of DNA, which was diluted 1:3 in the PCR reactions. We used PCR primers designed for the MA lines.

Sequence Analyses

We assembled all the generated sequences independently for each MA line and strain using SeqMan (DNASTAR, Inc.). All the trace files were visually inspected. Each single nucleotide polymorphism or indel in the MA lines was called if two independent PCR reactions verified the change. MA line base substitutions were marked as heteroplasmic if there was any evidence of the original progenitor base present in the chromatogram data. The frequencies of the mutant bases were estimated using the comparative peak height approach applied previously for the analysis of *D. melanogaster* MA line mtDNA divergence (Haag-Liautard et al. 2008). In all cases, the correction was carried out on both strands, and the frequencies averaged. All the assembled mitochondrial genomes were aligned using MUSCLE (Edgar 2004). The protein-coding genes of *P. pacificus* mtDNA were assigned according to the open-reading frames in SeqBuilder (DNASTAR, Inc.), set for the invertebrate mitochondrial codon table, and by multiple alignments using *C. elegans* and *Ascaris suum* mitochondrial genes for determining the initiation and termination codons. Nineteen of the 22 tRNA genes were identified by tRNAscan-SE 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy 1997) using the nematode mitochondria as source, strict search mode, and cove score

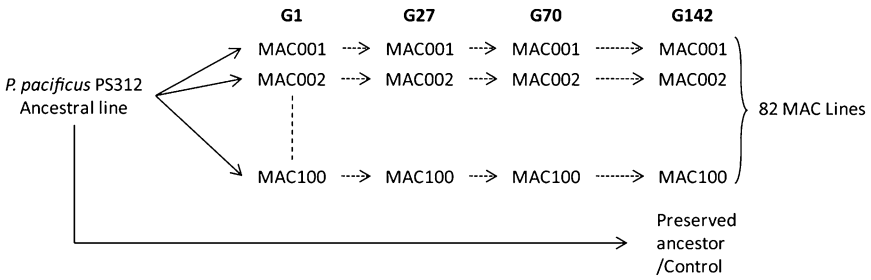


Fig. 1. Schematic representation of the experimental set up for the MA lines. 100 progeny from a single virgin hermaphrodite were used to start the experiment. At generations G1, G27, G70, and after the final generation G142, worms were frozen down to preserve the genomes of the mutated lines. A total of 82 lines survived the experiment.

cutoff of 20. With the cove score of 42.6, the $tRNA^{Tyr}$ has the highest cove score and the ones with the lowest scores are the $tRNA^{Arg}$ with 23.1 and $tRNA^{Sup}$ with 20.6. The secondary structures were determined in tRNAscan-SE. The remaining tRNA genes ($tRNA^{Ser(UCN)}$, $tRNA^{Ser(AGN)}$, and $tRNA^{Ile}$) were found based on alignments with the expected regions of the *C. elegans* mtDNA. The two rRNA genes were detected according to alignments with the *C. elegans* rRNA genes. The analyses of the pairwise comparison for the protein-coding genes in the intraspecific studies were performed using DnaSP v.5.10.01 (Librado and Rozas 2009), with the genetic code set to correspond to nematode mtDNA.

Mutation Rate Estimates

Mutation rates (per site per generation) were calculated as $\mu = m/(Lnt)$, and standard errors as $[\mu/(Lnt)]^{1/2}$, where m is the number of mutations, L is the number of lines, n is the number of base pairs analyzed per line, and t is the time in generations, as previously described (Denver et al. 2000). Heteroplasmy was inferred from the peak height and the frequencies calculated following equation (1) from Haag-Liautard et al. 2008. Taking this into account, we use equation (2) from Haag-Liautard et al. 2008 to estimate the overall mutation rate. We consider all mutations equally likely because the available number of mutations obtained in the MA line experiments does not yet justify to select any of the differential mutational models over any other model.

TMRCA Estimations

We used BEAST v1.6.1 (Drummond and Rambaut 2007) to estimate divergence times from the coding region of mtDNA data set for *P. pacificus* strains. We analyzed the data in three ways: 1) by third codon position, 2) partitioned according to codon position, and 3) by the entire coding region. The third codon data set consisted of 3,417 sites. For the partitioned data set, we created two alignments, one consisting of only the first and second codon positions (6,834 sites) and another one containing only the third codon position (3,417 sites). This is a proxy for the codon-based models that incorporate information about the genetic code (Shapiro et al. 2006). For the entire coding region, we used all 10,251 sites. For all alignments, we used the Bayesian Inference Criterion (Kass and Raftery 1995) to estimate the best-fitting model in jModelTest 0.1.1 (Guindon and Gascuel 2003; Posada 2008). All calculations were made assuming a strict molecular clock on all branches.

The taxon sets and implicitly the nodes for the divergence calculations were selected based on previously available information from two mitochondrial genes *nad6* and *nad4L* (Herrmann et al. 2010). For the divergence estimations, we allowed BEAST to infer topology of the trees. We included two speciation process models, birth–death (Gernhard 2008) and pure Yule process. We used Bayes factor as implemented in Tracer v1.5 (Drummond and Rambaut 2007) to select the best-fitting model under smoothed marginal likelihood estimate and with 100 bootstrap replicates (Suchard et al. 2001). In all cases, the birth–death was preferred by the Bayes factor analysis. Runs were initiated on random starting trees. A total of five independent runs of 10 million generations each were performed locally, sampling every 1000th generation, for each of the data sets. Convergence was assessed with Tracer v1.5. After discarding the first 1,000 samples as burn-in, the trees and parameter estimates from the five runs were combined using LogCombiner v1.6.1. The results were considered reliable once the effective sampling size of all parameters was above 100. Using TreeAnnotator v1.6.1 (Drummond and Rambaut 2007), the samples from the posterior were summarized on the maximum credibility tree, with the posterior probability limit set to 0.5 and summarizing mean node heights. These were visualized using FigTree v1.3.1 (Drummond and Rambaut 2007). Means and 95% higher posterior densities of age estimates were obtained from the combined outputs using Tracer v1.5.

Results

MA Lines

To provide the basis for mutation rate and divergence time estimates in *P. pacificus*, we generated MA lines in the strain PS312 from Pasadena (California, USA) (Sommer et al. 1996). This strain has been selected because it has been characterized in great detail and has been used as "wild-type" strain for all the developmental biology work carried out in the lab (Zheng et al. 2005; Tian et al. 2008). In addition, *P. pacificus* PS312 was sequenced with 9× coverage, and transcriptome and proteome date are available (Dieterich et al. 2008; Borchert et al. 2010). We initiated 100 MA lines of which 82 survived the 142 generations of single progeny propagation, giving a survival rate of 87.3% per 100 generations (fig. 1). In *C. elegans*, 87.8% of the MA lines survived for 100 generations, indicating a similar survival rate as observed in *P. pacificus* (Denver et al. 2000). From now on we refer to the MA lines of *P. pacificus* California as mutation accumulation lines California (MAC lines).

Mitochondrial Genome of *P. pacificus*

The mitochondrial genome of *P. pacificus* is 15,954 bp in size and is similar to other known nematode mitochondrial genomes (table 1 and fig. 2). It contains 12 protein-coding genes, two rRNAs, and 22 predicted tRNAs, ranging from 53 bp ($tRNA^{Ser(UCN)}$) to 63 bp ($tRNA^{Lys}$) in length. All genes are transcribed unidirectionally from the coding strand, which has an asymmetrical nucleotide composition of 44.0% T, 32.2% A, 14.9% G, and 8.9% C. The high AT composition is also reflected at the level of the individual genes, where *nad2* is 82.8% AT, whereas *cox1* is only 70.2%, these being the extremes encountered in the mtDNA of *P. pacificus*. Almost one-sixth of the genome is noncoding, comprising two large regions, the D-loop of 2,222 bp (76.33% AT) and a second region of 425 bp (75.76% AT), situated between $tRNA^{Asp}$ and $tRNA^{Gly}$. Other small intergenic regions are found, the biggest being the 66 bp region,

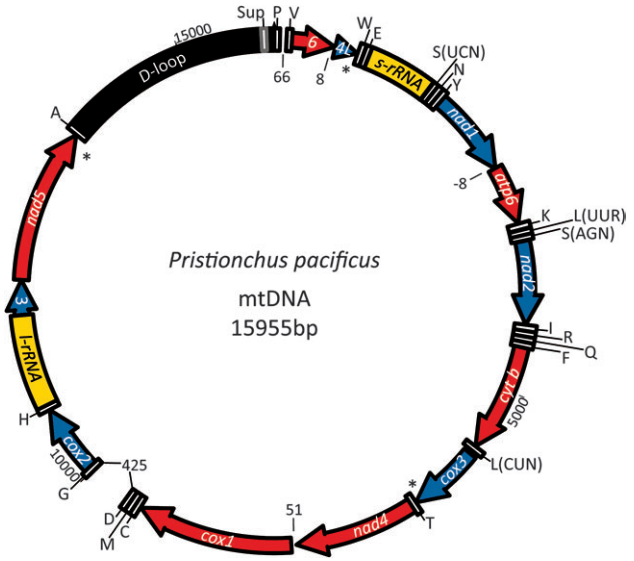


Fig. 2. Overview of the organization of the circular mtDNA of *Pristionchus pacificus*. The genes are colored according to the reading frame with red, green, and blue representing the first, second, and third reading frame, respectively, and s-rRNA and l-rRNA are represented in yellow. The tRNA genes, called by their amino acid symbol (Sup: Suppressor tRNA), are represented with white rectangles. Numbers at the inner periphery of the ring represent noncoding or overlapping nucleotides between neighboring genes. Numbers at the outer periphery indicate the length in base pairs. The incomplete termination codons are symbolized by the asterisk (*); 6, 4L, and 3 represent the genes *nad6*, *nad4L*, and *nad3*, respectively.

between $tRNA^{Pro}$ and $tRNA^{Val}$ and a 51 bp region between *nad4* and *cox1* (fig. 2). In *P. pacificus*, only 17.9% of the D-loop sequence consists of repeated sequences, contrary to *C. elegans*, where 55.4% of the D-loop sequence consists of different types of repeats (Okimoto et al. 1992). Specifically, in the case of *P. pacificus*, there is a run of 20 AT dinucleotides repeated in tandem. In addition, there are two repeats consisting each of 68, 27, and 12 nt, respectively, and two inverted repeats of 22 and 15 nt, respectively.

The mtDNA of *P. pacificus* has two characteristic features. First, the intergenic region situated between $tRNA^{Pro}$ and $tRNA^{Val}$, where Blast searches do not identify any homologous sequences. It is formed by a series of T:A and A:T homopolymeric stretches with a maximum of ten repeated nucleotides, but no microsatellite-like (di-, tri-, and tetranucleotide repeat) sequences have been identified. Second, the presence of a suppressor tRNA for the codon UAA, situated within the D-loop (fig. 2). The predicted folding of this suppressor tRNA, based on the tRNAscan-SE 1.21 software (Lowe and Eddy 1997), suggests this tRNA to have the typical truncated shape of nematode tRNAs (supplementary fig. S1, Supplementary Material online), although Blast search fails to identify any homologous sequence. The origin of this tRNA is unknown, but part of its sequence can be explained by two tandemly repeated sequences of 14 nt followed by two incomplete repeats of the same kind. UAA encodes a STOP codon in all the protein-coding genes in the *P. pacificus* mitochondrial genome, except for those genes that finish with an incomplete codon of a single U (*nad4L* or *nad5*) or a UA dinucleotide (*nad2* and *cox3*), that are possibly completed after polyadenylation (Ojala et al. 1981). As initiation codons, four of the six known possible alternatives are used in *P. pacificus*: AUU (*nad4L*, *nad1*, *atp6*, *cox3*, *nad3*, and *nad5*), AUA (*nad6*, *cyt b*, and *cox1*), UUG (*nad2* and *nad4*), and AUC (*cox2*).

The coding region of the mtDNA of *P. pacificus* consists of 3,417 codons, 45.3% are composed only of the nucleotides T and A, and in 73.9% of the codons contain at least one T or A. The most frequently used codon is UUU (11.7%), whereas the most abundant amino acid is leucine (14.5%). Of all the codons specifying leucine, 70% are UUA, the only leucine codon to contain only U and A. The least frequently used amino acid is arginine (0.9%). In general, the analysis of the relative synonymous codon usage shows that within each family, the codons with the highest content of A and U are preferred in the *P. pacificus* mitochondrial genome. Three codons are not used at all: GCG (alanine), CUC (leucine), and CGC (arginine), which are the most CG-rich codons of the families.

Mutation Rates and Spectrum

We made a direct estimate of the mutation rate in the mitochondrial genome in *P. pacificus* MA lines by sequencing 13,500 bp in 82 MA lines at 142nd generation. Among the 1,107,000 bp sequenced, we observed 12 mutations (table 3) resulting in a total mutation rate $\mu_{total} = 7.6 \times 10^{-8}$ ($\pm 2.2 \times 10^{-8}$) per site per generation.

Seven of these mutations were substitutions (table 3A), yielding a direct estimate of the mitochondrial mutation rate for base substitutions $\mu_{bs} = 4.5 \times 10^{-8}$ ($\pm 1.7 \times 10^{-8}$) per site per generation. This rate is lower than (but not significantly different from) the equivalent data for *C. elegans* of 9.7×10^{-8} per site per generation (Denver et al. 2000), *C. briggsae* (7.2×10^{-8} or 1.1×10^{-7} per site per generation depending of the strain used in the experiment) (Howe et al. 2010), and the mutation rate reported for *D. melanogaster* (mean mutation rate 6.2×10^{-8} per site per generation) (Haag-Liautard et al. 2008). The seven base substitutions found in the MA lines of *P. pacificus* occurred at different positions in the genome and in seven different lines, which suggests that there is no mutational hot spot in the nucleotides analyzed and that none of the lines is particularly predisposed to mitochondrial mutations. The mutation rate calculated for all mutations except the ones associated with repetitive sequence is 3.2×10^{-8} per site per generation (table 3A and B), which gives a mean number of mutations per line of 0.0609. Assuming a Poisson distribution, the expected number of lines with 0, 1, or 2 mutations is 77, 5, and 0, respectively. These values are almost identical to the observed numbers, where the number of lines with 0, 1, or 2 mutations are 71, 10, and 1, respectively, supporting the fact that there are no mutational hot spots, nor contamination. We found five transitions and two transversions (table 3), a ratio that agrees with mtDNA evolution, as observed at the intra- and interspecies level and characterized by a strong bias toward transitions (A:T to G:C and C:G to T:A) (Denver et al. 2000).

Table 3. Mutations in the Mitochondrial Genomes of the Mutation Accumulation Lines.

Line	Position	Mutation	Gene	Effect
A: Base substitution mutations				
MAC55[a]	1,631	A:T → G:C	*s-rRNA*	
MAC62[a]	2,834	C:G → T:A	*atp6*	Pro → Leu
MAC04	3,781	T:A → A:T	*nad2*	Leu → STOP
MAC30[a]	3,840	C:G → T:A	*nad2*	Silent
MAC24	4,587	A:T → G:C	*cyt b*	Silent
MAC06	11,636	G:C → T:A	*l-rRNA*	
MAC16	12,993	T:A → C:G	*nad5*	Silent
B: Indels				
MAC27	5,199	+1 T:A	*cyt b*	Trp → Cys + premature STOP
C: Homopolymer mutations				
MAC19	263	$(T:A)_8 \rightarrow (T:A)_7$	*nad6*	Leu → Trp + premature STOP
MAC78	6,336	$(T:A)_6 \rightarrow (T:A)_5$	*cox3*	Phe → Leu + premature STOP
MAC74	7,009	$(T:A)_6 \rightarrow (T:A)_5$	*nad4*	Phe → Leu + premature STOP
MAC16	12,351	$(T:A)_5 \rightarrow (T:A)_6$	*nad5*	Val → Cys + premature STOP

NOTE.—Positions denoted are with respect to the progenitor mtDNA sequence. MAC, MA lines California.
[a] Heteroplasmic mutations.

In addition to the seven base substitutions, we found five insertion–deletion (indel) mutations, yielding the indel mutation rate at $\mu_{indel} = 3.2 \times 10^{-8}$ ($\pm 1.4 \times 10^{-8}$), and four of them were associated with T:A homopolymeric nucleotide runs. Surprisingly, the longest homopolymers (9–11 bp T:A homopolymers) seem to be very stable in all the MA lines, regardless of their position in the mtDNA. The longest homopolymeric stretch mutated in *P. pacificus* is $(T:A)_8 \rightarrow (T:A)_7$, situated in the *nad6* gene. All the indel mutations are predicted to change the coding function of the respective genes (table 3B and C). Specifically, a single base pair insertion in a stretch of five T:A residues resulted in a premature STOP codon (one amino acid upstream the mutation) and reduced the NADH dehydrogenase subunit 5–87 amino acids lacking the conserved domain. Single base pair deletions occurred in a stretch of eight residues in the *nad6* gene and in a stretch of six residues each in *cox3* and *nad4*, all three of which resulted in premature STOP codon. One single base pair insertion occurred in a region without any type of repeat in the *cyt b* gene, resulting in an amino acid change and a premature STOP codon. In addition to these five mutations, one of the base substitution mutations (T:A to A:T in the line MAC04) resulted in a direct STOP mutation (table 3) in the *nad2* gene. Thus, 6 of the 12 mutations resulted in STOP codon mutations. The amount of viable mutations leading to premature STOP codons in *P. pacificus* MA lines is substantially higher than in *C. elegans* and *Drosophila* but similar than in *C. briggsae*. We speculate that the presence of the suppressor tRNA has influenced the spectrum of viable mutations in this organism (see Discussion).

The segregation of mitochondrial genomes allows for heteroplasmy, a phenomenon observed in multiple species. In *P. pacificus*, heteroplasmic sites were detected at three positions, all of them being single base pairs sites (table 3A). This differs from *C. briggsae*, where large heteroplasmic deletions have been found (Howe et al. 2010). The frequency of the heteroplasmic sites in *P. pacificus* were 0.4 for the MAC55 and MAC62, respectively and 0.3 for MAC30. Using these corrections, the total mutation rate is 6.4×10^{-8} per site per generation, a value that is in the error interval of our calculations.

Strains

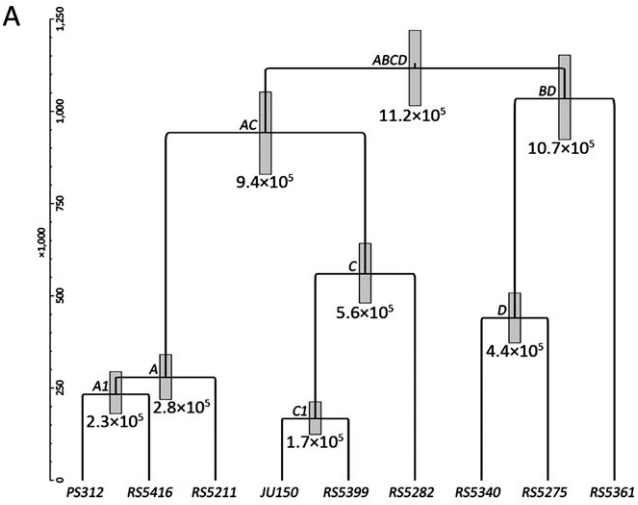
To study the genetic variation in the mtDNA of *P. pacificus* wild isolates, we have selected eight representative strains based on the available material from around the world to compare it to PS312 (table 2) (Herrmann et al. 2010). We amplified the entire mtDNA except for the D-loop for each strain with primers designed on the reference strain and then calculated the nucleotide differences relative to *P. pacificus* PS312. In the metazoan mtDNA, the D-loop is considered to be the most variable region, in addition to having a high AT content, making it very hard to amplify. We have been unable to amplify the D-loop from the other strains, even by using multiple alternative PCR primer sets. For all *P. pacificus* strains, the mtDNA coding region sums up to 10,281 bp, with very little variation in the AT content, at both the level of the entire coding region and the level of individual genes. The codon usage is similar between the tested strains with only a few exceptions: GCG (arginine) is used in all the strains except RS5361 and RS5211, the strain RS5340 is the only one that uses the codon CTC (leucine), and RS5275 is the only one that uses CGC (arginine). The genes *nad4L*, *nad2*, *cox3*, and *nad5* end with incomplete STOP codons in all the strains indicating that this feature of the *P. pacificus* mtDNA is conserved between wild isolates.

The comparison of synonymous and nonsynonymous changes between the different strains of *P. pacificus* (table 4) suggests that the strain RS5416 from the island of La Réunion is most closely related to PS312 with 165 (1.6%) synonymous and 30 (0.3%) nonsynonymous changes. The most distantly related strain is RS5399, also from La Réunion, with 482 (4.7%) synonymous and 45 (0.4%) nonsynonymous changes. These findings support the original observation of an unusually high genetic diversity of *P. pacificus* strains found on the island of La Réunion

(Herrmann et al. 2010). More general, these findings provide strong evidence for purifying selection with a magnitude similar to that reported for *C. elegans* and *C. briggsae* natural isolates (Howe and Denver 2008).

To obtain first insight into minimal divergence time estimates in *P. pacificus*, the information on the genetic variation in the mtDNA of wild isolates can be related to the mutation rate estimates from the MA line experiment. For the Bayesian analysis of the *P. pacificus* strains, we considered the first and second codon positions separately from the third codon position. We analyzed the data set in multiple ways and compare divergence time estimates obtained from three different procedures (see Materials and Methods). In a first approach, we only considered the third codon position, assuming complete neutrality (Cutter 2008). Considering the near absence of selection in the MA lines, we have used the overall mutation rate of 7.6×10^{-8} per site per generation for this analysis. In a second model, we partitioned the data set into two alignments composed of the first and second codon position together and, separately from that, the third codon position. This approach has the advantage of not eliminating the neutral mutations at the first and second codon positions. In a third model, we considered the entire coding region of the mtDNA.

The analysis at the third codon position yielded a point estimate of the time to the most recent common ancestor (TMRCA) for PS312 and its closest relative, RS5416 from La Réunion, of 2.9×10^5 generations (fig. 3). The most recent split occurred between the strain JU150 for Madagascar and RS5399 also from La Réunion, 2.3×10^5 generations ago. The divergence times of the other nodes show that the deepest analyzed divergence between the different strains of *P. pacificus* was 1.2×10^6 generations ago. Using the entire coding region or data partitioned by codon position does not significantly change the divergence time estimates (fig. 3*B*). For example, estimates based on the second model predict the divergence between PS312 and RS5416 to be 2.3×10^5 generations and the deepest split among all strains to be 1.1×10^6. Taken together, the divergence time estimates under these three models do not change significantly resulting in consistent relationships among strains and suggest a TMRCA at 10^5–10^6 generations (fig. 3*B*).



B

Dataset	A1 G×10⁵	A G×10⁵	AC G×10⁵	C1 G×10⁵	C G×10⁵	D G×10⁵	BD G×10⁵	ABCD G×10⁵
	Time to the Most Recent Common Ancestor							
I) 3rd codon position								
	2.9	3.4	10.7	2.3	6.0	4.4	10.9	12.0
95% HPD lower	2.5	3.0	9.4	1.9	5.2	3.9	9.3	11.0
95% HPD upper	3.4	4.0	12.5	2.8	6.9	5.1	12.4	13.0
II) Partitioned by codon position								
	2.3	**2.8**	**9.4**	**1.7**	**5.6**	**4.4**	**10.7**	**11.2**
95% HPD lower	1.8	2.2	8.3	1.2	4.8	3.7	9.3	10.1
95% HPD upper	2.9	3.4	1.1	2.1	6.4	5.1	12.2	12.2
III) Coding								
	1.2	1.4	4.3	0.9	2.4	1.8	4.5	4.7
95% HPD lower	1.0	1.2	3.7	0.7	2.1	1.5	3.9	4.3
95% HPD upper	1.4	1.6	5.0	1.1	2.7	2.0	5.0	5.1

Fig. 3. Phylogenetic relationship of *Pristionchus pacificus* isolates and time to the most recent common ancestor. (*A*) Phylogenetic relationship of the nine *P. pacificus* isolates based on the coding sequence of the mtDNA. The scale indicates the time in generations. The letters in the phylogram indicate the nodes for divergence calculations. (*B*) TMRCA for all eight nodes in generations. Calculations are based on 1) the third codon position, 2) the data set partitioned by codon position, and 3) the entire coding region. In bold are numbers represented in *A*.

Discussion

In this study, we provide mutation rate estimates of the mitochondrial genome of *P. pacificus* after 142 generations of inbreeding in 82 MA lines and compare these data with the genetic variation of the mtDNA of nine *P. pacificus* strains. The *P. pacificus* mitochondrial genome is 15,954 bp in size and is typical for nematodes, the difference in size resulting from the length of the noncoding regions (table 2). Almost one-sixth of the genome is noncoding, comprising two large and a number of small intergenic regions, from which, of additional importance, is the 51 bp region between *nad4* and *cox1*. This seems to be homologous to a region that in *C. elegans* has 109 bp, in *A. suum*, 117 bp (Okimoto et al. 1992), and in *Steinernema carpocapsae*, 44 bp (Montiel et al. 2006), located in the same boundaries in all species. Also, it can be folded in a hairpin structure with a run of four Ts in the loop. Such hairpin structures have been found in the noncoding regions of mtDNAs of other species, and in human and mouse, it has been shown to be the initiation sites of the synthesis of the second (L) strand of the molecule (Chang et al. 1985).

Although *P. pacificus* contains all known mitochondrial genes, the presence of a suppressor tRNA for the codon UAA is unusual. The presence of such a suppressor tRNA has most likely influenced the spectrum of observable mutations in the MA line experiments. Of the 12 obtained

Table 4. Pairwise Differences Between the Strains and Genes of *Pristionchus pacificus* Used in This Study.

Genes in *P. pacificus*	Length (bp)	Nonsynonymous/Synonymous Differences Relative to PS312							
	PS312	RS5416	RS5211	JU150	RS5399	RS5282	RS5340	RS5275	RS5361
nad6	435	3/8	3/9	4/19	5/19	9/19	8/20	5/21	6/19
nad4L[a]	232	1/3	1/3	1/10	2/9	0/8	0/7	0/9	3/12
nad1	876	1/12	1/11	0/21	1/24	3/41	2/38	2/42	4/42
atp6	600	1/5	0/2	2/7	1/6	2/22	0/21	0/16	0/17
nad2[a]	845	3/13	5/14	6/32	7/32	4/51	4/47	5/49	4/44
cyt b	1,104	2/13	5/14	5/34	5/44	6/34	5/38	3/45	6/40
cox3[a]	767	0/16	0/16	1/28	1/29	1/31	0/36	1/39	4/38
nad4	1,230	3/17	3/21	7/51	6/52	5/54	6/59	5/55	4/52
cox1	1,578	0/31	0/34	1/57	1/60	0/60	0/73	1/75	1/76
cox2	696	0/12	0/17	1/24	1/26	1/21	0/29	0/35	2/27
nad3	336	0/17	0/6	1/14	1/15	2/13	0/12	1/14	4/9
nad5[a]	1,582	7/28	5/29	5/95	7/95	5/97	8/96	6/93	7/106
Total	10,281	30/165	23/176	34/392	38/411	38/451	33/476	29/493	45/482
%AT[b]	75.93	75.86	75.84	75.62	75.62	75.85	75.92	75.97	75.39

[a] Incomplete STOP codon.
[b] Calculated only for the protein-coding regions.

mutations, six result in nonsense mutations, all of which can be suppressed by the suppressor tRNA. In *C. elegans* in comparison, similar MA line experiments resulted in the identification of 26 mutations after 214 generations in 74 MA lines, but only four of these mutations resulted in a premature STOP codon. However, given the small number of total mutations, it remains unclear if the values observed in *P. pacificus* and *C. elegans* are statistically different from one another. Considering the number of mutations, *P. pacificus* has nearly three times as many nonsense mutations than expected. From the *C. elegans* data set (Denver et al. 2000) of four nonsense mutations in a total of 26, only two nonsense mutations would have been expected for *P. pacificus*. In contrast, the number of expected and observed mutations in *P. pacificus* is the same when considering the number of nucleotides scanned; based on the four nonsense mutations in *C. elegans* from 10,428 bp in 74 lines, the expected number in *P. pacificus* from 13,500 bp in 82 lines would be six. Although the statistical significance of the different number of nonsense mutations remains open, we speculate that the *P. pacificus* nonsense mutations did not result in a lethal phenotype and were not subject to purifying selection. Thus, the presence of a suppressor tRNA apparently influenced the type and number of mutations that have been accumulated in *P. pacificus* MA lines. It should be noted however, that no premature STOP codons have been found in the mtDNA genes of the *P. pacificus* natural isolates, which argues for presumptive fitness affects under natural conditions.

The *P. pacificus* MA line–derived mutation rate is 7.6×10^{-8} ($\pm 2.2 \times 10^{-8}$) per site per generation, different than the values found in *C. elegans* (Denver et al. 2000) or in *C. briggsae* (Howe et al. 2010). However, the differences among these nematodes are not statistically significant. Recent studies indicated that mutation rates might vary according to the genetic background (Howe et al. 2010). Although this observation influences the estimation of divergence times, the real consequences cannot currently be evaluated. Future studies can reveal if features of the mitochondrial genomes of *C. elegans* and *P. pacificus* also hold true for nuclear genomes. One advantage of MA line experiments is that the resulting lines can be analyzed for multiple traits nearly indefinitely, making them a powerful tool for evolutionary investigations.

One major aim of this study was to relate genetic variation in wild isolates to the mutation rate estimates from MA lines in order to obtain first insight into minimal divergence times in *P. pacificus* (fig. 3). The divergence time estimates derived from the three tested models show consistent relationships among strains and suggest a minimal divergence of 10^5–10^6 generations. The number of generations that *P. pacificus* produces in nature each year depends on several biological and ecological parameters. From a theoretical perspective, this number might vary between 1 and 100 generations per year, with 100 generations representing the maximum number of generations that can be produced under optimal growth conditions when worms go through a direct life cycle without entering the arrested dauer stage (Sommer et al. 1996). In contrast, one generation per year represents the absolute minimum, which we have observed by studying the survival rate of *P. pacificus* dauer larvae (Mayer and Sommer 2011). Several considerations favor the hypothesis that the actual number of generations of *P. pacificus* per year is closer to our lower theoretical estimate than to the maximum number of generations that is observed in the laboratory. First, our field studies indicate that *Pristionchus* nematodes occur on scarab beetles exclusively in the arrested dauer stage (Weller et al. 2010). Second, we have no evidence that *Pristionchus* would reproduce on the living beetle; rather, they wait for the beetle's death in the dauer stage. Therefore, we assume that there is only one *Pristionchus* generation on a particular beetle (Herrmann et al. 2006a, 2007). Third, most scarab beetle species have very slow generation times, for example, the *P. pacificus* host *E. orientalis* in the USA and Japan has an annual life cycle (Herrmann et al. 2007). Fourth, our limited data on the soil reservoir of *Pristionchus* nematodes, which are based on studies in German

forests, also indicate that the majority of *Pristionchus* nematodes found in nature are in the dauer stage (Weller AM, Sommer RJ, unpublished observation). Finally, although our experimental work so far only provided evidence for a necromenic association with scarab beetles, we can formally not rule out that under certain conditions, *P. pacificus* might have a phoretic relationship with these beetles.

Assuming a number of one to ten generations per year, the data provided in this study would suggest that the divergence between the nine analyzed strains of *P. pacificus* is somewhere between 10^4–10^5 years. Considering a fairly complete sampling of major groups of haplotypes within the species, these estimates might represent the most basal divergence within *P. pacificus*. This number is consistent with the available molecular phylogeny (Mayer et al. 2007, 2009) as well as with the spectrum of beetle hosts, and the biogeography of islands that *P. pacificus* is found on (Herrmann et al. 2006a, 2006b, 2007, 2010). Further insight and confirmation of the natural history of the *P. pacificus* require the identification and characterization of a closely related sister species, an endeavor that is still ongoing.

Supplementary Material

Supplementary table S1 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Werner E. Mayer, Katy Morgan, Richard Neher, and Erik Ragsdale for critically reading the manuscript and for fruitful discussions, Dan Bumbarger, Angela McGaughran, Martin Mechelke, Michael Matschiner, Robbie Rae, Amit Sinha, and Adrian Streit for fruitful discussions. We would like to thank T. Eirich and S. Maeck for preparing the plates during the MA lines experiment and H. Haussmann for freezing the MAC lines. Accession codes: The nucleotide sequence reported here has been deposited at GenBank with the accession number JF414117. This work was funded by the Max-Planck Society.

References

Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* 154:181–192.

Avise JC. 1998. The history and purview of phylogeography: a personal reflection. *Mol Ecol*. 7:371–379.

Baer CF, Shaw F, Steding C, et al. (11 authors). 2005. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc Natl Acad Sci U S A*. 102:5785–5790.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.

Blaxter ML, De Ley P, Garey JR, et al. (12 co-authors) 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature* 392:71–75.

Borchert N, Dieterich C, Krug K, Schütz W, Jung S, Nordheim A, Sommer RJ, Macek B. 2010. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res*. 20:837–846.

Chang DD, Hauswirth WW, Clayton DA. 1985. Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *EMBO J*. 4:1559–1567.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 10:195–205.

Cutter AD. 2008. Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol*. 25:778–786.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679–682.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* 289:2342–2344.

Dieterich C, Clifton SW, Schuster L, et al. (17 co-authors). 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet*. 40:1193–1198.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5:113.

Gernhard T. 2008. The conditioned reconstructed process. *J Theor Biol*. 253:769–778.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol*. 6:1706–1714.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445:82–85.

He Y, Jones J, Armstrong M, Lamberti F, Moens M. 2005. The mitochondrial genome of *Xiphinema americanum sensu stricto* (Nematoda: Enoplea): considerable economization in the length and structural features of encoded genes. *J Mol Evol*. 61:819–833.

Herrmann M, Kienle S, Rochat J, Mayer WE, Sommer RJ. 2010. Haplotype diversity of the nematode *Pristionchus pacificus* on Réunion in the Indian Ocean suggests multiple independent invasions. *Biol J Linn Soc*. 100:170–179.

Herrmann M, Mayer WE, Hong R, Kienle S, Minasaki R, Sommer RJ. 2007. The nematode *Pristionchus pacificus* (Nematoda: Diplogastridae) is associated with the oriental beetle *Exomala orientalis* (Coleoptera: Scarabaeidae) in Japan. *Zool Sci*. 24:883–889.

Herrmann M, Mayer WE, Sommer RJ. 2006a. Nematodes of the genus *Pristionchus* are closely associated with scarab beetles and the Colorado potato beetle in Western Europe. *Zoology*. 109:96–108.

Herrmann M, Mayer WE, Sommer RJ. 2006b. Sex, bugs and Haldane's rule: the nematode genus *Pristionchus* in the United States. *Front Zool*. 3:14.

Hong RL, Sommer RJ. 2006. *Pristionchus pacificus*: a well-rounded nematode. *BioEssays* 28:651–659.

Hong RL, Witte H, Sommer RJ. 2008. Natural variation in *P. pacificus* insect pheromone attraction involves the protein kinase EGL-4. *Proc Natl Acad Sci U S A*. 105:7779–7784.

Howe DK, Baer CF, Denver DR. 2010. High rate of large deletions in *Caenorhabditis briggsae* mitochondrial genome mutation processes. *Genome Biol Evol*. 2010:29–38.

Howe DK, Denver DR. 2008. Muller's Ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. *BMC Evol Biol*. 8:62.
Hu M, Chilton NB, Gasser RB. 2002. The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *Int J Parasitol*. 32: 145–158.
Hu M, Chilton NB, Gasser RB. 2003. The mitochondrial genome of *Strongyloides stercoralis* (Nematoda) - idiosyncratic gene order and evolutionary implications. *Int J Parasitol*. 33: 1393–1408.
Hyman BC, Azevedo JLB. 1996. Similar evolutionary patterning nematode mitochondrial genes among repeated and single copy. *Mol Biol Evol*. 13:221–232.
Jacob JE, Vanholme B, Van Leeuwen T, Gheysen G. 2009. A unique genetic code change in the mitochondrial genome of the parasitic nematode *Radopholus similis*. *BMC Res Notes*. 2:192.
Kass R, Raftery A. 1995. Bayes factors. *J Am Stat Assoc*. 90:773–795.
Keddie EM, Higazi T, Unnasch TR. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Mol Biochem Parasitol*. 95:111–127.
Lavrov DV, Brown WM. 2001. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* 157:621–637.
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 Res*. 25:955–964.
Lynch M. 2007. The origin of genome architecture. Sunderland (MA): Sinauer.
Lynch M. 2010. Evolution of the mutation rate. *Trends Genet*. 26:345–352.
Lynch M, Sung W, Morris K, et al. (11 co-authors). 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*. 105:9272–9277.
Mayer MG, Sommer RJ. 2011. Natural variation in *Pristionchus pacificus* dauer formation reveals cross-preference rather than self-preference of nematode dauer pheromones. *Proc R Soc B Biol Sci*. Advance access published February 9, 2011,doi:10.1098/rspb.2010.2760.
Mayer WE, Herrmann M, Sommer RJ. 2007. Phylogeny of the nematode genus *Pristionchus* and implications for biodiversity, biogeography and the evolution of hermaphroditism. *BMC Evol Biol*. 7:104.
Mayer WE, Herrmann M, Sommer RJ. 2009. Molecular phylogeny of beetle associated diplogastrid nematodes suggests host switching rather than nematode-beetle coevolution. *BMC Evol Biol*. 9:212.
Megen H van, Elsen S van den, Holterman M, Karssen G, Mooyman P, Bongers T, Holovachov O, Bakker J, Helder J. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11:927–950.
Montiel R, Lucena MA, Medeiros J, Simões N. 2006. The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae*: insights into nematode mitochondrial DNA evolution and phylogeny. *J Mol Evol*. 62:211–25.
Mukai T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19.
Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474.
Okimoto R, Chamberlin HM, Macfarlane JL, Wolstenholme DR. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. *Nucleic Acids Res*. 19:1619–1626.
Okimoto R, Macfarlane JL, Clary D, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130:471–498.
Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in. *Arabidopsis thaliana*. *Science* 327:92–94.
Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25:1253–1256.
Schlager B, Wang X, Braach G, Sommer RJ. 2009. Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode model *Pristionchus pacificus*. *Genesis* 47:300–304.
Seyfert AL, Cristescu ME, Frisse L, Schaack S, Thomas WK, Lynch M. 2008. The rate and spectrum of microsatellite mutation in Caenorhabditis elegans and Daphnia pulex. *Genetics* 178: 2113–2121.
Shapiro B, Rambaut A, Drummond AJ. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol*. 23:7–9.
Sommer RJ. 2009. The future of evo-devo: model systems and evolutionary theory. *Nat Rev Genet*. 10:416–422.
Sommer RJ, Carta LK, Kim S-Y, Sternberg PW. 1996. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda, Diplogastridae). *Fundam Appl Nematol*. 19:511–521.
Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol*. 18:1001–1013.
Tian H, Schlager B, Xiao H, Sommer RJ. 2008. Wnt signaling by differentially expressed Wnt ligands induces vulva development in *Pristionchus pacificus*. *Curr Biol*. 18:142–146.
Vassilieva LL, Lynch M. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* 151: 119–129.
Weller AM, Mayer WE, Rae R, Sommer RJ. 2010. Quantitative assessment of the nematode fauna present on Geotrupes dung beetles reveals species-rich communities with a heterogeneous distribution. *J Parasitol*. 96:525–531.
Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol*. 141:173–216.
Zauner H, Sommer RJ. 2007. Evolution of robustness in the signaling network of Pristionchus vulva development. *Proc Natl Acad Sci U S A*. 104:10086–10091.
Zheng M, Messerschmidt D, Jungblut B, Sommer RJ. 2005. Conservation and diversification of Wnt signaling function during the evolution of nematode vulva development. *Nat Genet*. 37:300–304.
Zouros E, Freeman KR, Ball AO, Pogson GH. 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel Mytilus. *Nature* 359:412–414.

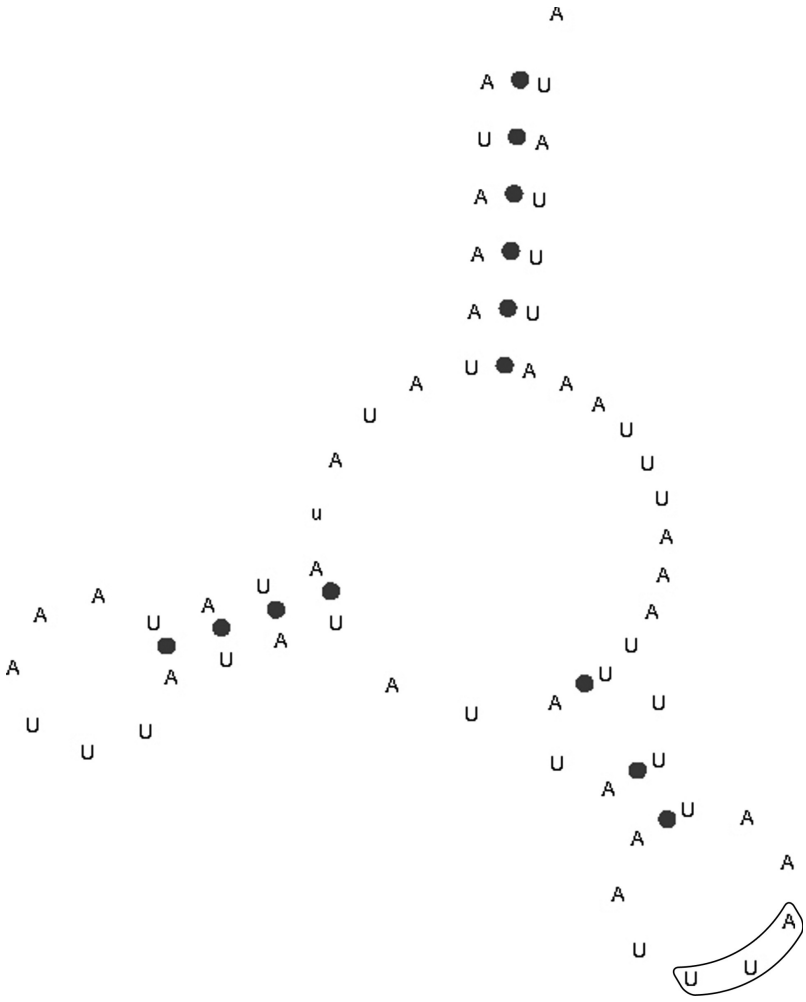
Supplementary material for

Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*

Supplementary Table 1. PCR primers used for *P. pacificus* mtDNA amplification

PCR fragment	Primer name and sequence
ND6-4L	ND6-4L_2 GAACTATAAATCGGCCGAAGG
	ND6-4LRev AACCTTTTCTGCGAAAAAGAAA
12SrRNA	12SrRNAFor AGGTTGTCTACAGCATTAAGAGAAA
	12SrRNARev TTGTTTAAGGATACAACCCTTTGA
ND1	ND1For TTGAAAACATGAAGCGAGTACAA
	ND1Rev AACAGATCTCAAAGGTAGTCTAGCA
ATP6	ATP6_F AGTTCTTATCCACGTTATCGTTATG
	ATP6_R TTTGCGCTTAAAACAAATGC
ND2	ND2_F CCAAGGTTTAGAATTAAGAATAGGAG
	ND2_R AACATGATATATTTCACCAGAAAAGTT
ND3	ND3_F AAAGTGTTTAGTACGAAAGGAACAA
	ND3_R CCCACTCTACAAAATACAAACTCA
CYB	Gap31For TTATCAGCATTAACATTTAGATATTGG
	COX3_R1 ATTGTGACCCAAAGAACCAA
COX3	CYB_1 CTTCAAGATTGGTGAAATTAAATAAA
	ND4_R1 TGAGAACCATAACCTAAAATTATTACC
ND4	COX3_1 ATAACCACCATTTAGGGTTTGA
	COX1_R1 TGTTCAAACGTGGAAACCTT
COX1	ND4_1 TCAGAATTTTTAGTTGTAACAAATGG
	COX2_R1 AATAATTCTCCAAATTGGTATTCAA
COX2	COX1_4 GGTTTTCCGCGTAAATATATGG
	16SrRNA_R1 CTCACGCTAAGACTGCCATT
16SrRNA	COX2_1 TTGGATGCTATAAGAGGTGTTTT
	Gap31R1 AAAACTGACGTGAAACCAAACA
ND5	ND5_F3 GATGTCATTTATTGTTGGTGGTT
	ND5_R1 TGACTGAAGAGCAGTAAAAGTTGG

Supplementary Figure 1



Tandem-Repeat Patterns and Mutation Rates in Microsatellites of the Nematode Model Organism *Pristionchus pacificus*

Ruxandra I. Molnar, Hanh Witte, Iris Dinkelacker, Laure Villate, and Ralf J. Sommer[1]
Max-Planck Institute for Developmental Biology, Department for Evolutionary Biology, Spemannstrasse 37, D-72076 Tübingen, Germany

ABSTRACT Modern evolutionary biology requires integrative approaches that combine life history, population structure, ecology, and development. The nematode *Pristionchus pacificus* has been established as a model system in which these aspects can be studied in one organism. *P. pacificus* has well-developed genetic, genomic, and transgenic tools and its ecologic association with scarab beetles is well described. A recent study provided first mutation rate estimates based on mitochondrial genome sequencing and mutation accumulation line experiments that help resolve rather ancient evolutionary branches. Here, we analyzed the tandem-repeat pattern and studied spontaneous mutation rates for microsatellite markers by using the previously generated mutation accumulation lines. We found that 0.59%–3.83% of the genome is composed of short tandem repeats. We developed 41 microsatellite markers, randomly chosen throughout the genome and analyzed them in 82 mutation accumulation lines after 142 generations. A total of 31 mutations were identified in these lines. There was a strong correlation between allele size and mutation rate in *P. pacificus*, similar to *Caenorhabditis elegans*. In contrast to *C. elegans*, however, there is no evidence for a bias toward multistep mutations. The mutation spectrum of microsatellite loci in *P. pacificus* shows more insertions than deletions, indicating a tendency toward lengthening, a process that might have contributed to the increase in genome size. The mutation rates obtained for individual microsatellite markers provide guidelines for divergence time estimates that can be applied in *P. pacificus* next-generation sequencing approaches of wild isolates.

KEYWORDS
mutation accumulation lines
Pristionchus pacificus
microsatellite markers
tandem repeats
Caenorhabditis elegans

The nematode *Pristionchus pacificus* is a model organism increasingly used for integrative approaches in evolution biology, through interdisciplinary studies in evo-devo, population genetics, and ecology (Hong and Sommer 2006; Sommer 2009). *P. pacificus* has a generation time of 4 days in standard laboratory cultures (Sommer *et al.* 1996), well-developed tools for forward and reverse genetic analysis, DNA-mediated transformation (Schlager *et al.* 2009), and a fully sequenced genome (Dieterich *et al.* 2008). *Pristionchus* nematodes are unique among model organisms in their well-described necromenic association with scarab beetles, *e.g.*, *P. pacificus* has been found on *Exomala orientalis* in Japan and on *Oryctes borbonicus* on La Réunion Island in the Indian Ocean (Herrmann *et al.* 2007, 2010; Morgan *et al.* 2012). The *Pristionchus*–beetle association represents a robust platform for the isolation and characterization of new *Pristionchus* isolates on a global scale. Close to 30 *Pristionchus* species and more than 400 *P. pacificus* strains have been isolated between 2004 and 2011 in worldwide samplings, and a molecular phylogenetic framework has been generated (Mayer *et al.* 2007, 2009; Morgan *et al.* 2012).

Integrative approaches in evolutionary biology require a life history perspective. Specifically, robust evolutionary analyses depend on an understanding of the mutation patterns of different regions of the genomes (Lynch 2007). In a recent study, we used a mutation accumulation (MA) line approach to evaluate the pattern of mutations and to estimate the mutation rates of the mitochondrial genome of *P. pacificus* (Molnar *et al.* 2011). These can be used to resolve ancient

doi: 10.1534/g3.112.003129
Manuscript received May 12, 2012; accepted for publication June 22, 2012

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.112.003129/-/DC1.
[1]Corresponding author: Max-Planck Institute for Developmental Biology, Department for Evolutionary Biology, Spemannstrasse 37, D-72076 Tübingen, Germany. E-mail: ralf.sommer@tuebingen.mpg.de

Table 1 Observed repeat loci in the *P. pacificus* genome

Repeat unit	Perfect Repeats		Imperfect Repeats[a]	
	TRF–Strict	TRF–Loose	TRF–Strict	TRF–Loose
Total	730	Not all sites are mentioned		
Dinucleotides				
AC	21	1599	29	434
AG	265	10,785	1636	7510
AT	97	1541	316	647
CG	0	549	7	54
Total	383	14,474	1988	8645
Trinucleotides				
AAC	32	4155	632	1869
AAG	3	9336	606	5762
AAT	57	6437	629	3236
ACC	2	1093	8	1159
ACG	2	1858	21	399
ACT	3	1072	20	241
AGC	3	2762	359	1382
AGG	3	7247	573	4161
ATC	28	4461	79	1088
CCG	0	718	23	193
Total	133	39,139	2,950	19,490
Tetranucleotides				
AAAC	0	444	23	416
AAAG	1	902	38	1381
AAAT	4	2250	175	2921
AACC	0	4	0	36
AACG	0	174	0	70
AACT	2	86	7	63
AAGC	0	127	0	54
AAGG	0	1000	63	1616
AAGT	0	73	9	97
AATC	8	1457	63	1213
AATG	29	1426	122	1130
AATT	3	1222	835	2822
ACAG	0	239	70	205
ACAT	1	175	11	117
ACCC	0	68	218	26
ACCG	0	51	1	85
ACCT	0	82	3	57
ACGC	0	48	0	41
ACGG	0	46	1	35
ACGT	0	7	0	2
ACTC	1	309	3	223
ACTG	0	125	3	71
AGAT	2	325	8	240
AGCG	0	240	1	263
AGCT	0	8	0	0
AGGC	0	52	1	27
AGGG	1	995	0	2681
ATCC	0	330	0	283
ATCG	0	476	3	208
ATGC	0	44	0	19
CAGC	1	30	0	14
CCCG	0	38	0	54
CCGG	0	12	0	330
Total	53	12,922	1837	17,219
Pentanucleotides[b]				
AAAAG	2	80	16	878
AAAAT	1	312	86	1615

(continued)

Table 1, *continued*

Repeat unit	Perfect Repeats		Imperfect Repeats[a]	
	TRF–Strict	TRF–Loose	TRF–Strict	TRF–Loose
AAAGG	1	50	11	469
AAAGT	12	107	9	136
AAATC	4	101	11	236
AAGAG	1	316	201	2088
AAGGG	2	189	58	1100
AATAC	4	103	10	194
AATTC	55	334	48	453
ACATC	2	32	1	80
ACTCT	1	11	1	73
AGAGG	0	122	113	1274
AGGGG	0	139	167	1816
Total	85	Not all sites are mentioned		
Hexanucleotides[b]				
AAAAAC	1	56	66	338
AAAAAG	1	31	88	505
AAAAAT	0	46	70	790
AAAAGT	1	5	5	59
AAAATC	1	4	2	48
AAAGGT	2	158	5	226
AAATTC	11	18	5	59
AAATTG	20	47	11	95
AACAAT	2	36	203	457
AAGAGG	1	42	127	618
AAGCCT	8	43	68	101
AAGTAT	3	21	12	27
AATAAG	6	24	101	207
AATATC	0	32	2	33
AATCTG	1	3	1	17
AATTAC	6	27	9	55
ACACGC	1	1	2	12
ACCAGG	1	1	0	2
ACTCGC	1	1	210	892
AGAGGG	0	78	6	14
AGCCGG	1	2	0	9
AGAGTC	1	1	10	45
ATCTGT	2	50	7	45
ATCGTC	2	5	6	34
ATCTTC	1	14	83	212
ATGATT	2	4	9	57
Total	76	Not all sites are mentioned		

[a] Shown only if perfect loci were found.
[b] Shown only if perfect loci were found with the TRF-strict method or more than 30 perfect loci were found with TRF-loose method.

evolutionary branches, whereas more recent evolutionary events are better studied using microsatellite regions. To most effectively use microsatellites for divergence estimation, we need to understand their mutation pattern and the factors that affect their mutation rate.

Microsatellites are DNA sequences composed of short units, no more than 6 bp long, found as tandem repeats throughout the genomes of most eukaryotic and prokaryotic organisms (Hancock 1999). They are ubiquitously but nonrandomly distributed in protein-coding and -noncoding regions (Toth *et al.* 2000). Their highly polymorphic nature made microsatellites the markers of choice in population genetics (Jarne and Lagoda 1996). Several mechanisms have been suggested to explain the high mutation rate of microsatellites, including errors during recombination, unequal crossing-over, and polymerase slippage during DNA replication (Schlötterer and Tautz 1992) or repair (Strand *et al.* 1993). Individual microsatellites are described as having a life cycle—they are born, they grow, and they die (Chambers and MacAvoy 2000).

Estimations of the rate and pattern of microsatellite mutations are usually indirect, based on allele frequency distributions (Chakraborty *et al.* 1997; Primmer and Ellegren 1998) or phylogenetic analyses (Jin *et al.* 1996; Dettman and Taylor 2004). The first studies aiming to understand the mutation mechanisms of microsatellites have been made possible by direct observations of mutations based on artificial constructs with expressed microsatellite sequences within bacterial and fungal systems (Levinson and Gutman 1987; Strand *et al.* 1993). Direct estimates of the microsatellite mutation rates are also derived from MA line experiments conducted in *Drosophila melanogaster* (Schug *et al.* 1997), *Caenorhabditis elegans*, *Daphnia pulex* (Seyfert *et al.* 2008), and *Arabidopsis thaliana* (Marriage *et al.* 2009). Under ideal conditions, MA line-based mutation rate estimates can be combined with genomic analysis of natural isolates of a given species and close relatives to provide robust divergence time estimates (Molnar *et al.* 2011).

Here, we evaluate the genomic composition of microsatellites for *P. pacificus* and make use of MA lines to provide robust estimates of

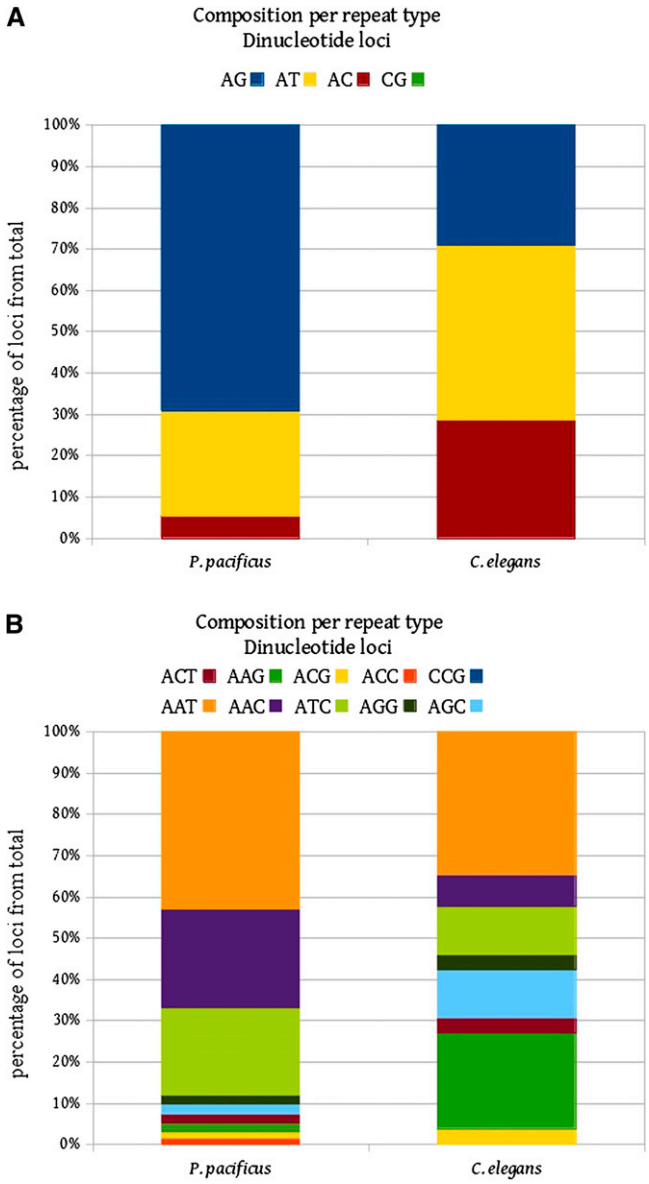


Figure 1 Composition per repeat type: (A) dinucleotide loci, (B) trinucleotide loci.

the rate and size spectra of microsatellite mutations. We found that 0.59%–3.83% of the genome is composed of short tandem repeats. By analyzing 82 MA lines after 142 generations, we found a total of 31 mutations in these markers. There is a correlation between allele size and mutation rate, but no bias toward multistep mutations. We use these findings to suggest general guidelines for the selection of microsatellite markers in future genome-wide association studies of the evolutionary history of *P. pacificus*.

MATERIALS AND METHODS

MA lines

The propagation of the MA lines has been described in detail previously (Molnar *et al.* 2011). To summarize, 100 MA lines were initiated from the F3 descendants of a single, inbred, wild-type *P. pacificus* PS312 laboratory strain. MA lines were propagated by a single, randomly chosen offspring from the middle of the reproduction period. Worms were cultured at 20° on nematode growth medium seeded with *Escherichia coli* as a food source, as originally described for *P. pacificus* (Sommer *et al.* 1996). Backup cultures were kept for two generations at 15° to prevent the accidental loss of lines. From the original 100 MA lines, 82 lines survived the 142 generations. Because a single homozygous progenitor started all the lines, a single random offspring began each generation and because *P. pacificus* is a self-fertilizing hermaphrodite, the variation between the strains is caused by the accumulation of mutations.

Microsatellite pattern

Microsatellite loci were identified in the 'Freeze 1' assembly of the *P. pacificus* genome (available at www.pristionchus.org) using Tandem Repeats Finder (TRF) (Benson 1999) with two different sets of parameters: threshold alignment score 20 and alignment weights {2,7,7} ({match,mismatch,indels}) for the TRF-loose method and threshold alignment score 50 and alignment weights {2,3,5} for the TRF-strict method (Leclercq *et al.* 2007). The values for the weights can be 3, 5, and 7, with 3 being more permissive and 7 more restrictive. Leclercq *et al.* have previously shown that both the weights and threshold criteria influence the composition of the microsatellites detected (see also supporting information, Figure S1 for the differential screening results from the present study). Specifically, Leclercq *et al.* have shown that increasing TRF alignment score allows the detection of smaller and more perfect microsatellites, whereas decreasing the TRF weights allows for longer and more imperfect microsatellites to be detected. We, therefore, have created two data sets of microsatellite loci. Both datasets include perfect and imperfect repeats. The frequency of microsatellites was calculated as (number of microsatellite loci)/(number of nucleotides in megabases), considering the assembled genome of *P. pacificus* of 169 Mb, and loci of at least three repeat units. For a comparison base line, we screened the *C. elegans* genome with the same methods as for *P. pacificus*.

Microsatellite markers for molecular analysis

We randomly chose 32 microsatellites and developed suitable primers (Table S1). These markers cover a range of three to 57 repeat units and have a percentage match score between 51 and 100. To ensure the evaluation of the mutation rate for the long microsatellites present in the genome, we chose nine additional markers for perfect tri-, tetra-, penta-, and hexanucleotide that had pattern repeat count greater than 30 (except the hexanucleotide locus that had the pattern repeat count of 17 but an overall length of more than 100 bp). We had 41 markers in total, represented by 19 perfect and 22 imperfect repeat loci.

DNA extraction and amplification

For each MA line, genomic DNA was prepared from two full plates of worms, using worm lysis buffer (50 mM KCl; 10 mM Tris-HCl, pH 8.3; 2.5 mM $MgCl_2$; 0.45% NP-40; 0.45% Tween-20; 5 μg/mL proteinase K). The suspension was incubated for 2 hr at 65°, followed by inactivation of the proteinase K at 95° for 10 min. All forward primers had an M13 tail (5′-CACGACGTTGTAAAACGAC-3′) attached at the 5′ end, labeled with 6-FAM, VIC, NED, or PET (Applied Biosystems) for genotyping (Table S1). We performed individual polymerase chain reactions in 20 μL of final volume for each marker, using an annealing temperature of 55°.

Genotyping

Genotyping was performed on an ABI 3730*xl* using ABI Genemapper version 4.0 (Applied Biosystems) analysis software with the internal size standard GS500LIZ or GS1200LIZ. Markers M01 to M47 were multiplexed before genotyping. Markers M74 to M88 were labeled

Table 2 Allelic mutation rate estimates per generation

Marker	Chr.	Repeat	Percent Match	No. Mutations	Magnitude of Mutation (No. Lines)	Mutation Rate[a]
M21	II	$(GGGCAC)_{11}$	51	0		
M01	I	$(CT)_{55}$	55	0		
M15	II	$(TC)_{57}$	55	0		
M41	V	$(TCT)_{26}$	59	0		
M07	I	$(TTG)_{18}$	64	0		
M43	V	$(TTAA)_{15}$	69	0		
M29	III	$(TTAA)_{10}$	72	0		
M16	II	$(TGA)_{15}$	73	0		
M35	IV	$(CTCC)_{20}$	73	0		
M42	V	$(AATT)_{9}$	75	0		
M13	I	$(CTTAAC)_{6}$	78	0		
M08	I	$(TAAT)_{3}$	81	0		
M47	X	$(GT)_{7}$	84	0		
M06	I	$(AAC)_{16}$	85	0		
M34	IV	$(AT)_{17}$	85	1	+1(1)	(9.2×10^{-5})
M26	III	$(TTA)_{29}$	85	0		
M04	I	$(CAA)_{11}$	87	0		
M22	II	$(GAATAA)_{5}$	88	0		
M45	V	$(GAGAG)_{3}$	91	0		
M11	I	$(TTCTT)_{3}$	92	0		
M33	IV	$(AG)_{17}$	93	0		
M14	I	$(GAATAA)_{4}$	95	0		
M28	III	$(TGG)_{3}$	100	0		
M18	II	$(TCT)_{3}$	100	0		
M17	II	$(GTT)_{3}$	100	1	−1(1)	(8.9×10^{-5})
M19	II	$(TCGA)_{3}$	100	0		
M38	IV	$(AAGCCT)_{6}$	100	0		
M03	I	$(TC)_{7}$	100	0		
M05	I	$(AGT)_{8}$	100	0		
M02	I	$(GA)_{13}$	100	0		
M25	III	$(AG)_{14}$	100	0		
M46	X	$(AG)_{15}$	100	0		
M84	III	$(CTCTTC)_{17}$	100	1	+1(1)	(8.9×10^{-5})
M74	V	$(CAA)_{33}$	100	2	−1(1), +1(1)	2.3×10^{-4}
M77	X	$(ACAT)_{35}$	100	1	+1(1)	(9.6×10^{-5})
M78	I/V	$(ACAT)_{36}$	100	3	+1(2), -9(1)	2.8×10^{-4}
M79	I	$(TTAG)_{40}$	100	4	+1(4)	3.6×10^{-4}
M82	X	$(TGAAT)_{43}$	100	6	−3(1), -1(1), +1(2), +2(1), +3(1)	7.5×10^{-4}
M88	?	$(TC)_{59}$	100	3	−1(2), -2(1)	2.7×10^{-4}
M80	?	$(CAGC)_{64}$	100	2	+1(2)	2.1×10^{-4}
M83	?	$(TTCAA)_{64}$	100	7	+1(7)	6.5×10^{-4}

[a] Mutation rate is calculated per locus per generation (the number of lines assessed for each marker may differ). Mutation rates given in parentheses are based on single mutation events.

with 6-FAM and assessed individually. Mutations were detected with Genemapper version 4.0 (Applied Biosystems) by comparison to the progenitor of the MA lines and verified using independent DNA amplification and genotyping. All the assessed loci were homozygous.

Mutation rate estimates

The mutation rates (per allele per generation) were calculated using the formula from Seyfert *et al.* (2008): $\mu = -[\ln(1-n/l)]/t$, where n is the number of mutations, t is the number of generations, and l is the number of lines. Note that the number of lines assessed may differ slightly between the markers.

RESULTS

Tandem repeat pattern in *P. pacificus*

We screened the genome of *P. pacificus* for short tandem repeats consisting of di-, tri-, tetra-, penta-, and hexanucleotides with at least three repeat units, using the Tandem Repeats Finder software (Benson 1999). Data presented below always refer to duplex DNA, even if we show only the sequence of the repeated motif on one strand for simplicity, *i.e.*, notations like $(AC)_n$ and $(AC)_n$:$(GT)_n$ are equivalent. The two sets of parameters, involving different thresholds and distinguishing between perfect and imperfect repeats (see *Materials and Methods*), yielded very different results. Specifically, in the genome of *P. pacificus*, the TRF-loose method counted 70,543 perfect loci, whereas the TRF-strict method identified only 730 loci. From the 730 perfect loci identified by the TRF-strict method, dinucleotide repeats are by far the most common repeat type (383/730). Of the four possible unique dimer combinations (AC, AG, AT, CG), three are present within the genome, with AG repeats representing the greatest number (265/383 loci; Table 1). In contrast, perfect CG repeats are not found in the *P. pacificus* genome. In comparison with the *C. elegans* genome, the dinucleotide repeats have a different repeat unit distribution, but the perfect CG repeats are also missing (Figure 1A).

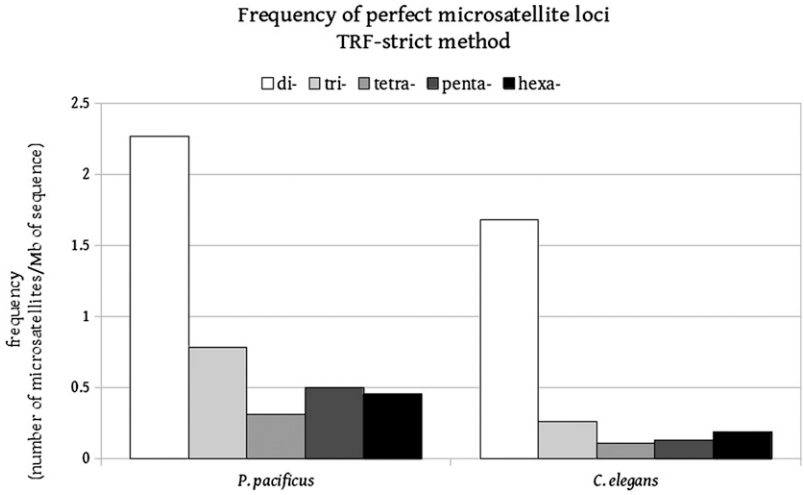


Figure 2 Frequency of perfect microsatellite loci, TRF-strict method.

A total of 133 loci were found representing the 10 possible unique trimer combinations. Trimers rich in AT are the most common, with $[(A/T)_2, X_1]$ combinations accounting for 92% of the instances found (Table 1). Again, one possible combination is not found, and this is the pure CCG repeat (Table 1, Figure 1B). There are 33 unique tetramer combinations possible, 11 of which have been found in the *P. pacificus* genome. 90% of the tetramer repeats are represented by $[(A/T)_3,X_1]$ type of repeats. Most of the repeat units with more than 50% CG-content ($[(A/T)_1(C/G)_3]$ and $[(A/T)_2(C/G)_2]$) were missing, including $[(C/G)_4]$. Similarly, 90% of the pentamers and 72% of the hexamers are represented by AT-rich motifs ($[(A/T)_4,X_1$ and $[(A/T)_5, X_1]$, respectively; Table 1). The 'nematode' telomeric repeat motif, $(TTAGGC)_n$ (Niedermaier and Moritz 2000), is found to be the most abundant among the hexamer repeats (8 pure loci in TRF-strict dataset—the most in terms of number of loci per repeat type and the longest repeats per locus), but no functional conclusion can be drawn.

The TRF-loose method yielded in general the same trend for the different types of repeats, with AT-rich loci representing more than 50% of the total number of loci found for each repeat unit size. CG-rich pentamer and hexamer repeats are few or missing (Table 1).

Mutation rates

We randomly selected 32 markers from the loci found with the TRF-loose method (the larger dataset). To these we added nine markers selected based on the total repeat count of more than 30 to have the longest loci represented in the analysis. Among the 41 loci assayed in the 82 MA lines, we found 31 mutation events at 11 loci (Table 2). At four of these 11 loci, only a single mutation was observed, which is insufficient for deriving proper mutation frequencies. Therefore, most conclusions listed below are based on markers with more than one mutation. Several general trends can be observed in the mutation patterns. First, M17 was the only locus smaller than 17 units that contained a mutation. In general, microsatellite loci that contain more than 30 repeat units show more mutations than the shorter ones, an observation similar to the *C. elegans* MA line-based analysis of microsatellites mutation rates (Seyfert *et al.* 2008). Second, 26 of the 31 mutations involved gain or loss of a single repeat unit, with insertions outnumbering deletions (21 insertions and 5 deletions). Only five of the 31 mutations involved a change of more than one repeat unit, and only one large deletion (−9 repeats at the M78) occurred at a locus with more than 30 repeat units. Third, most of the mutations were found in perfect, long microsatellite loci (29 of 31 mutations; Table 2). Fourth, in the overall MA pattern, deletions are unable to balance out the insertions (8 *vs.* 23 repeat units). This observation suggests that the microsatellite loci have a tendency toward lengthening. Further analysis shows that seven of the analyzed loci (M34, M77, M79, M80, M82, M83, and M84) show accumulated growth, two (M17, M78) show accumulated decrease in length, and two (M74, M88) show no accumulated change in size. Finally, the mutation rate per locus per generation in *P. pacificus* ranges from 8.9×10^{-5} to 7.5×10^{-4} for those markers where mutations occurred.

DISCUSSION

This is the first analysis of mutational processes and mutation rate estimates in the nuclear genome of *P. pacificus* because previous knowledge is based solely on the mitochondrial genome (Molnar *et al.* 2011). We have analyzed the tandem repeat pattern of the *P. pacificus* genome and studied the spontaneous mutation rates for microsatellite markers. From the mutation patterns and mutation rates obtained for individual microsatellite markers, we provide guidelines for the properties of microsatellite markers to be useful for divergence time estimates in future genome-wide sequencing projects (to follow in this section).

P. pacificus and *C. elegans* belong to the same nematode clade but they are only distantly related, representing members of different nematode families (Dieterich *et al.* 2008). Sequence turnover over these evolutionary distances resulted in unrelated microsatellite patterns in these two species. Therefore, the microsatellites are not homologous and cannot be directly compared, which unfortunately prevents the usage of statistical methods in a meaningful manner. The average overall AT content for the *P. pacificus* genome is 58% (Dieterich *et al.* 2008); therefore, we would expect AT dinucleotide repeats to be more common, followed by AC and AG in approximate equal numbers, and fewer CG loci. This pattern is, however, not followed by the dimer repeats composition found in *P. pacificus* genome, arguing against the expectation of cryptic simplicity. In contrast, the tri- to hexanucleotide repeat loci do follow this expectation, with AT-rich repeats being more abundant than the others. The most striking finding, however, is the absence or near absence of the CG loci. Direct and indirect observations tend to support the stepwise mutation model at microsatellite loci (Schlötterer and Tautz 1992; Weber and Wong 1993; Thuillet *et al.* 2002), by which their sequence is altered by addition or deletion of one repeat at a time. An alternative is the model according to which the sequence of microsatellites can be

altered by large deletions, due to secondary structures that certain types of repeats can form (Di Rienzo *et al.* 1994). The absence of perfect CG dimer repeats but the presence of impure and CG-rich loci might support the latter model of microsatellite evolution.

A comparison with the *C. elegans* microsatellite dataset reveals that *P. pacificus* has an overall greater frequency of perfect microsatellite loci, although the dinucleotide repeats dominate the landscape in both genomes (Figure 2), a finding that has also been made by Castagnone-Sereno *et al.* (2010) using different algorithms. In a second step, we evaluated the mutation rate at microsatellite loci ranging from di- to hexanucleotide repeats, randomly chosen in the noncoding genome of *P. pacificus*. The random choice allows us to avoid a bias by assaying only certain types of repeats. However, it does not allow us to make decisions of how the mutation rate is influenced by the repeat unit size, nucleotide composition, or the overall length of the locus. Although the loci have been chosen randomly, eight of 10 dinucleotide loci are of the type $(AG)_n$. This correlates with the general composition of the genome of *P. pacificus*, which has more AG repeats than other dinucleotide repeats.

The mutation patterns observed for *P. pacificus* in this study support the idea that mutational processes are length dependent. Specifically, large loci have, on average, more mutations than the small ones (three mutations in a $(X)_3$ locus *vs.* seven mutations in a $(X)_{64}$ locus). A second general trend, supported by the *P. pacificus* data, is that imperfect loci are less prone to accumulate mutations than the perfect ones. From 22 loci with diverse percentage of match and with a wide range of repeat unit size tested, only one showed a mutation (M34; Table 2). A comparison of the mutation rates at microsatellite loci with the same range of repeat number between *C. elegans* (Seyfert *et al.* 2008) and *P. pacificus* shows a similar effect of repeat number on mutation rates in both organisms.

A third major trend in the *P. pacificus* dataset is that the mutation process is upwardly biased in that loci tend to gain units more frequently that they lose units. Similar trends have been shown previously in other organisms (Primmer *et al.* 1996; Ellegren 2000). At the same time, long alleles tend to contract upon mutation (Harr and Schlötterer 2000; Xu *et al.* 2000). The mutations observed in *P. pacificus* are 21 insertions and 5 deletions, showing that microsatellites in *P. pacificus* have a tendency toward lengthening. It is interesting to note that the largest *P. pacificus* microsatellites detected in the genomes assembly are substantially smaller than the largest *C. elegans* microsatellites, which have repeat units greater than 68 (Seyfert *et al.* 2008). In *P. pacificus*, the loci M80 and M83, both show insertions, indicating that they are still in the growth phase. We speculate that the *P. pacificus* microsatellites, on average, are still in the expansion phase, a process that might have substantially contributed to the increase of the *P. pacificus* genome size relative to *C. elegans*. A final aspect of our analysis is that the *P. pacificus* genome shows no evidence for a bias toward multistep mutations. Specifically, all but five mutations are single-step insertions or deletions. This pattern is clearly distinct from what has been observed in *C. elegans*, indicating, again, the species and locus-specificity of the mutational processes.

The rate and pattern of mutations observed in the MA lines have implications for the use of microsatellites for inference of genetic history. It is critical to recognize that the evolutionary rate for a single locus will change with the size of the allele. Thus, choosing a microsatellite locus with the appropriate evolutionary rate to address a specific evolutionary time frame requires a careful consideration of allele size ranges. The data presented in this study provide guidelines for the selection of adequate markers for studying recent and ancient evolutionary branches of *P. pacificus*. Specifically, the absence of mutations in many short and/or imperfect loci, as well as the fact that for five markers only one mutation has been identified, do not allow us to use these results in deriving a mean mutation frequency. Interestingly however, all but one of the perfect repeats with more than 30 repeat units recovered multiple mutation events that resulted in a quite stable mutation rate of $2.5-7.5 \times 10^{-4}$ (Table 2). Therefore, we suggest to only use microsatellite markers with a minimal length of 30 repeat units in studies that aim to reconstruct the evolutionary history of wild isolates. Furthermore, we suggest that such studies should use an average mutation frequency of 5×10^{-4} given the relatively stable mutation frequencies obtained in this study. Taken together therefore, this study provides useful information for future genome-wide studies that investigate the evolutionary history of *P. pacificus*.

ACKNOWLEDGMENTS

We thank Amit Sinha, Werner Mayer, Katy Morgan, and Richard Neher for discussions and critically reading the manuscript; Christian Rödelsberger and Moritz Muschick for discussions on the microsatellite analysis; Sebastian Fröhler for running the initial TRF scan; Heike Haussmann for thawing the MA lines; and Tamara Eirich for preparing the NGM plates.

LITERATURE CITED

Benson, G., 1999 Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27: 573–580.

Castagnone-Sereno, P., E. Danchin, E. Deleury, T. T. Guillemaud, T. Malausa, and P. Abad, 2010 Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne incognita*. BMC Genomics 11: 598.

Chakraborty, R., M. Kimmel, D. Stivers, L. Davison, and R. Deka, 1997 Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA 94: 1041–1046.

Chambers, G., and E. MacAvoy, 2000 Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology Part B 126: 455–476.

Dettman, J., and J. Taylor, 2004 Mutation and evolution of microsatellite loci in Neurospora. Genetics 168: 1231–1248.

Dieterich, C., S. W. Clifton, L. N. Schuster, A. Chinwalla, K. Delehaunty *et al.*, 2008 The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. 40: 1193–1198.

Di Rienzo, A., A. Peterson, J. Garza, A. Valdes, M. Slatkin *et al.*, 1994 Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA 91: 3166–3170.

Ellegren, H., 2000 Heterogeneous mutation processes in human microsatellite DNA sequences. Nat. Genet. 24: 400–402.

Hancock, J., 1999 Microsatellites and other simple sequences: genomic context and mutational mechanism, pp. 1–9 in *Microsatellites—Evolution and Applications*, edited by D. Goldstein, and C. Schlötterer. Oxford University Press, Oxford.

Harr, B., and C. Schlötterer, 2000 Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. Genetics 155: 1213–1220.

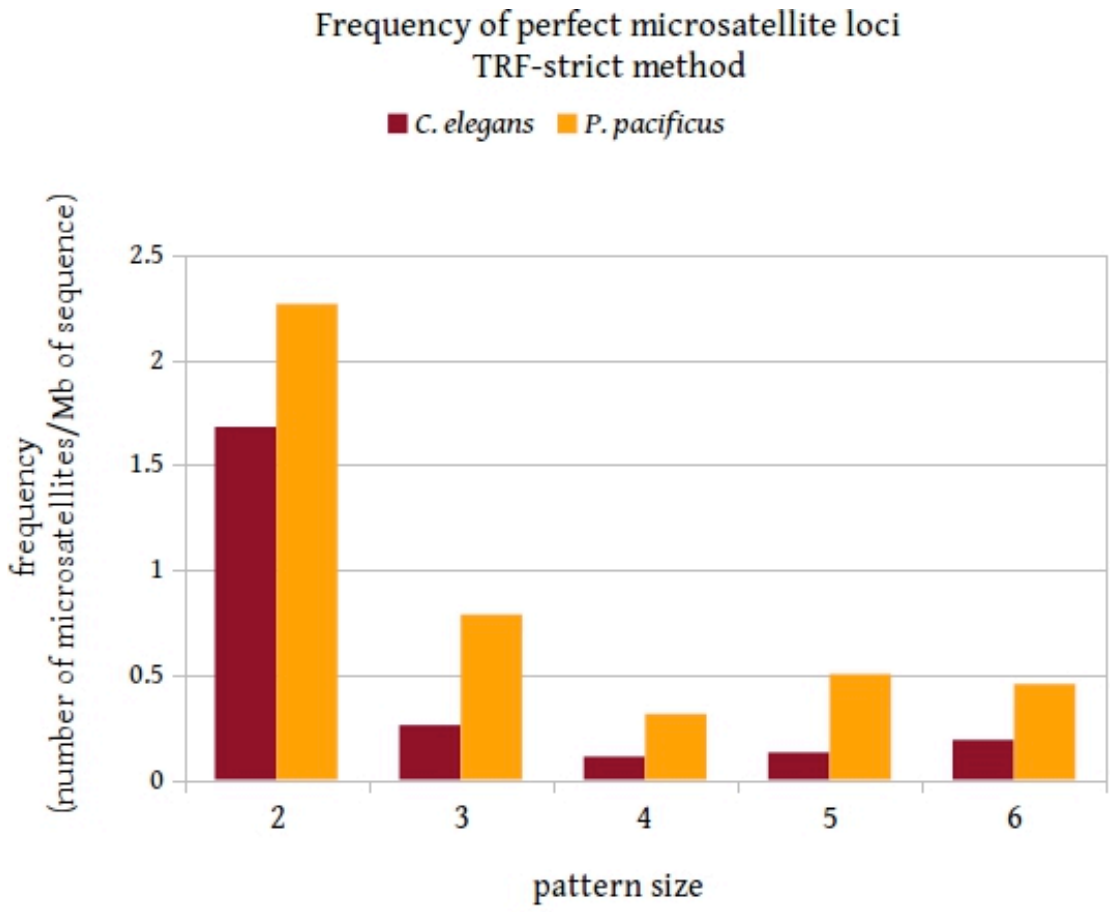
Herrmann, M., W. Mayer, R. Hong, S. Kienle, R. Minasaki *et al.*, 2007 The nematode *Pristionchus pacificus* (Nematoda: Diplogastridae) is associated with the oriental beetle *Exomala orientalis* (Coleoptera: Scarabaeidae) in Japan. Zoolog. Sci. 24: 883–889.

Herrmann, M., S. Kienle, J. Rochat, W. Mayer, and R. Sommer, 2010 Haplotype diversity of the nematode Pristionchus pacificus on Réunion in the Indian Ocean suggests multiple independent invasions. Biol. J. Linn. Soc. Lond. 100: 170–179.

Hong, R., and R. Sommer, 2006 *Pristionchus pacificus*: a well-rounded nematode. Bioessays 28: 651–659.

Jarne, P., and P. Lagoda, 1996 Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11: 424–429.
Jin, L., C. Macaubas, J. Hallmayer, A. Kimura, and E. Mignot, 1996 Mutation rate varies among alleles at a microsatellite locus : Phylogenetic evidence. Genetics 93: 15285–15288.
Leclercq, S., E. Rivals, and P. Jarne, 2007 Detecting microsatellites within genomes: significant variation among algorithms. BMC Bioinformatics 8: 125.
Levinson, G., and G. Gutman, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4: 203–221.
Lynch, M., 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, MA.
Marriage, T., S. Hudman, M. Mort, M. Orive, R. Shaw *et al.*, 2009 Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). Heredity 103: 310–317.
Mayer, W., M. Herrmann, and R. Sommer, 2007 Phylogeny of the nematode genus Pristionchus and implications for biodiversity, biogeography and the evolution of hermaphroditism. BMC Evol. Biol. 7: 104.
Mayer, W. E., M. Herrmann, and R. J. Sommer, 2009 Molecular phylogeny of beetle associated diplogastrid nematodes suggests host switching rather than nematode-beetle coevolution. BMC Evol. Biol. 9: 212.
Molnar, R. I., G. Bartelmes, I. Dinkelacker, H. Witte, and R. J. Sommer, 2011 Mutation rates and intraspecific divergence of the mitochondrial genome of *pristionchus pacificus*. Mol. Biol. Evol. 28: 2317–2326.
Morgan, K., A. McGaughran, L. Villate, M. Herrmann, H. Witte *et al.*, 2012 Multi locus analysis of *Pristionchus pacificus* on La Réunion Island reveals an evolutionary history shaped by multiple introductions, constrained dispersal events and rare out-crossing. Mol. Ecol. 21: 250–266.
Niedermaier, J., and K. B. Moritz, 2000 Organization and dynamics of satellite and telomere DNAs in *Ascaris*: implications for formation and breakdown of compound chromosomes. Chromosoma 109: 439–452.
Primmer, C., and H. Ellegren, 1998 Patterns of Molecular Evolution in Avian Microsatellites. Mol. Biol. Evol. 15: 997–1008.
Primmer, C., N. Saino, A. Møller, and H. Ellegren, 1996 Directional evolution in germline microsatellite mutations. Nat. Genet. 13: 391–393.
Schlager, B., X. Wang, G. Braach, and R. J. Sommer, 2009 Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. Genesis 47: 300–304.
Schlötterer, C., and D. Tautz, 1992 Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20: 211–215.
Schug, M., T. Mackay, and C. Aquadro, 1997 Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nat. Genet. 15: 99–102.
Seyfert, A., M. Cristescu, L. Frisse, S. Schaack, W. Thomas *et al.*, 2008 The Rate and Spectrum of Microsatellite Mutation in *Caenorhabditis elegans* and *Daphnia pulex*. Genetics 178: 2113–2121.
Sommer, R. J., 2009 The future of evo-devo model systems and evolutionary theory. Nat. Rev. Genet. 10: 416–421.
Sommer, R. J., L. Carta, S.-Y. Kim, and P. Sternberg, 1996 Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda : Neodiplogastridae). Fundam. Appl. Nematol. 19: 511–521.
Strand, M., T. Prolla, R. Liskay, and T. Petes, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274–276.
Thuillet, A.-C., D. Bru, J. David, P. Roumet, S. Santoni *et al.*, 2002 Direct Estimation of Mutation Rate for 10 Microsatellite Loci in Durum Wheat, *Triticum turgidum*. Mol. Biol. Evol. 19: 122–125.
Toth, G., Z. Gaspari, and J. Jurka, 2000 Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome Res. 10: 967–981.
Weber, J., and C. Wong, 1993 Mutation of human short tandem repeats. Hum. Mol. Genet. 2: 1123–1128.
Xu, X., M. Peng, and Z. Fang, 2000 The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. 24: 396–399.

Communicating editor: D. G. Moerman



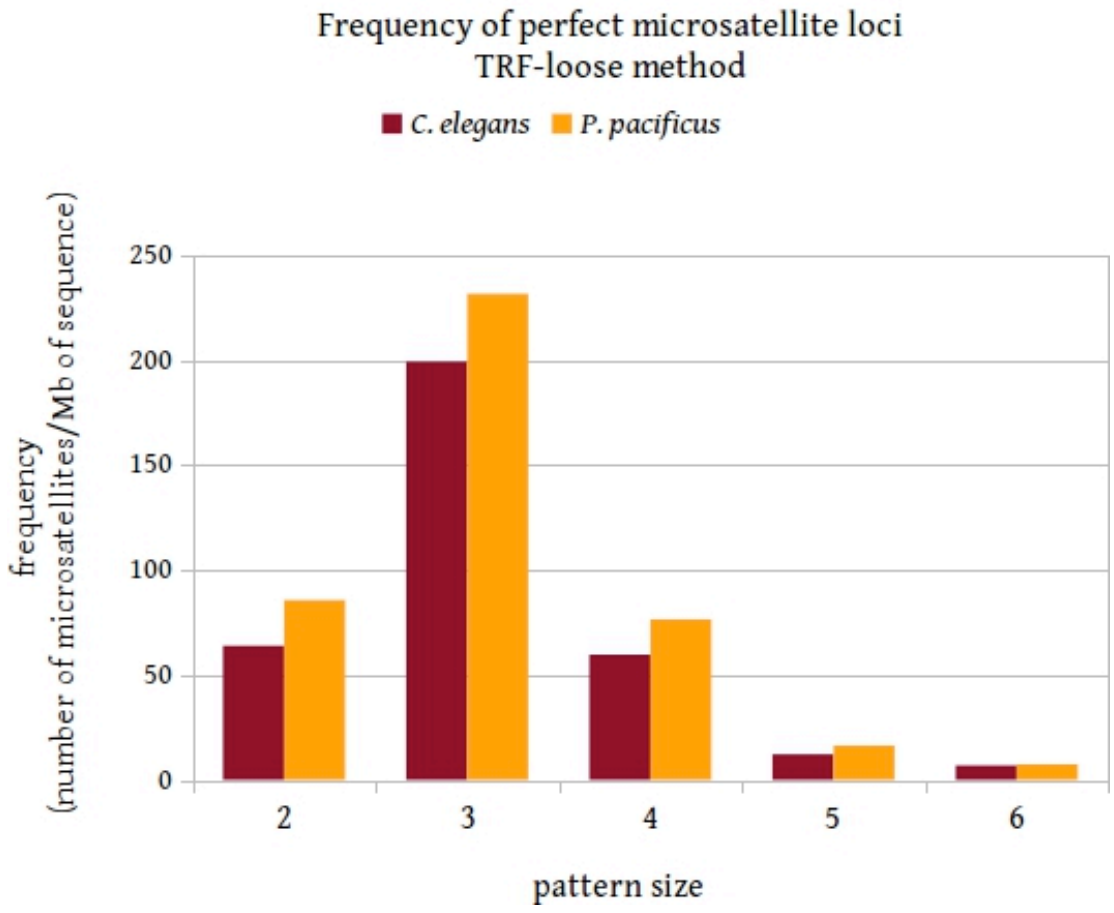


Figure S1 Differential screening results.

Table S1 Microsatellite loci assayed in *P. pacificus*

Marker (Dye)	Chr	Repeat	Percent match	Forward primer 5'- 3'/ Reverse primer 5'- 3'
			Randomly chosen markers	
M01 (6-FAM)	I	$(CT)_{55}$	55	AAAGCAATGCCAGAAGGAAA / TTCTTGCCGTACAAAGAATGA
M02 (NED)	I	$(GA)_{13}$	100	TTGGCTTTCTACAGCTAAATCG / GCCTGTGAAGGGGATATTGA
M03 (VIC)	I	$(TC)_{7}$	100	GTTTGTCAGCCCGATATGGT / GGGATGGGAAGGGAGAAAC
M04 (VIC)	I	$(CAA)_{11}$	87	GCTGGCATTGGAAAGAAAGT / AAGCCGATCAATCTCTGTCA
M05 (PET)	I	$(AGT)_{8}$	100	CGATTCCGTCGCTGTTTTC / AAGCCGATCAATCTCTGTCA
M06 (PET)	I	$(AAC)_{16}$	85	AGGCTCGTGTCGTTCATGTT / GCCAGCTACGGATGATTGAT
M07 (PET)	I	$(TTG)_{18}$	64	GCACAATCGGGTCTGAAAAT / TCCGGAACCTACAAAAGTGG
M08 (PET)	I	$(TAAT)_{3}$	81	GGTAAGTGCGCTGGAGAGAG / GCCCGAGGACACCTACTACA
M11 (6-FAM)	I	$(TTCTT)_{3}$	92	AACCGAATGTGGACGAAGAG / GGTCATCCCTTGTTTGTGCT
M13 (VIC)	I	$(CTTAAC)_{6}$	78	AAAGGCGAGGGTACAGTCCT / CGAGCAACTGAAATGAACGA
M14 (PET)	I	$(GAATAA)_{4}$	95	CTCCCTACCTCCCACATCAA / ATACTCGGTGGCCAGTATGC
M15 (VIC)	II	$(TC)_{57}$	55	CGGCTTCCTTCCTCTCTTCT / CCAACTCTCCCCTTTTCACA
M16 (PET)	II	$(TGA)_{15}$	73	GCCTGGACAAAATTGTCGTT / CGTTGTGGTGTGAGTTACGG
M17 (VIC)	II	$(GTT)_{3}$	100	AAGACCTGGGCATCAAACTG / ACGACCGAAACATCTTGGAC
M18 (VIC)	II	$(TCT)_{3}$	100	TCTAACTGGATGCCGGAACT / AGCGGTGTTAATGAGCGTTT
M19 (NED)	II	$(TCGA)_{3}$	100	ACATTGTCCATTCAGCTTCG / CCGAAAGAGAGACCATCTCC
M21 (6-FAM)	II	$(GGGCAC)_{11}$	51	GGAAAGGAGAGGGGAGTCTG / TCAGCCTCCTCAGGTAATGA
M22 (6-FAM)	II	$(GAATAA)_{5}$	88	ATTACGAGCTGGCCAATCAG / AAGTGCTACACTCGGTGCAA
M25 (NED)	III	$(AG)_{14}$	100	AGCAACGGCTCAATCAAAGT / GCAGTGACTCTTTCCGGTTT
M26 (VIC)	III	$(TTA)_{29}$	85	CTGATTGCCCGACAACATTC / CCAACGCATCACTATGGCTA
M28 (6-FAM)	III	$(TGG)_{3}$	100	GATTCGGTGGTGTAGTCGATG / TTCTGTGCATGACGAAGTTTCT
M29 (6-FAM)	III	$(TTAA)_{10}$	72	GTGTCGTGTGCCGACTAAGA / CTCTCCGGTTCCTCCTCTCT
M33 (6-FAM)	IV	$(AG)_{17}$	93	TGATTGACTGAGGGCTTTCC / ATGCCACCCTTCTGATTGAC
M34 (6-FAM)	IV	$(AT)_{17}$	85	ATTTGTAGGGGAAGGGGTTG / CATAAACGGTGACAGGCACA
M35 (6-FAM)	IV	$(CTCC)_{20}$	73	TTGGTCTTCTCCCCTTCTGA / TCTCCTTGTTTCCCCTCCTT
M38 (6-FAM)	IV	$(AAGCCT)_{6}$	100	CTCAATCAGGGAGGAACCAA / CCTCTCCCTCTCGACAACTG
M41 (NED)	V	$(TCT)_{26}$	59	CATCAATGAAACCCCATTCC / ATAACAGGCGCTGCTCTCTC
M42 (6-FAM)	V	$(AATT)_{9}$	75	GAAACACGAATGCCCACTCT / TCAGAGATGCCCGAGAGTTT
M43 (6-FAM)	V	$(TTAA)_{15}$	69	CTCCATCGCCTTCTTCTTTG / CCGGTCTATTCCTGGTCGTA
M45 (PET)	V	$(GAGAG)_{3}$	91	ATAGAAGAAGCGGTGGGTCA / CTCCACGTTCACTCGTCTCA
M46 (PET)	X	$(AG)_{15}$	100	CGCAATGAAGAGAACGAGGT / GAGCCGAGAAGTCCAGTGAG
M47 (6-FAM)	X	$(GT)_{7}$	84	GTCGGTCAAATATCGCCTGT / CGTCTGTCCCCGTTATCACT
			Tri- to hexanucleotide microsatellites with more than 30 repeat count	
M74 (6-FAM)	V	$(CAA)_{33}$	100	CCTCAGTCCCAACAGCAGAT / TTCCCAACACTTTGCATGAG
M77 (6-FAM)	X	$(ACAT)_{35}$	100	GGATCGTCCCTTTCGTCATA / CACGAAACACACAGCGTTCT
M78 (6-FAM)	I/V	$(ACAT)_{36}$	100	GGTCTGAGGGGTGTACGATG / TCATCATTGCCGAATTGTGT
M79 (6-FAM)	I	$(TTAG)_{40}$	100	AAAGTCGAGCGTGACAGACA / AATACGCATCTCGCTCTGGT
M80 (6-FAM)	?	$(CAGC)_{64}$	100	AAAGGGGTGAAATTTGCATT / GAGAAGTATGCAGGGCTGTTG

M82 (6-FAM)	X	$(TGAAT)_{43}$	100	TTGCAGGTTCGATACATTCTTC / GTTTCCCCCTGCAGGTCTAT
M83 (6-FAM)	?	$(TTCAA)_{64}$	100	GAAAATGTCGCCACGAAAAT / TGCAAACGTTACACAAACAGC
M84 (6-FAM)	III	$(CTCTTC)_{17}$	100	ACGATGATGATGTGCGTGAG / AATCGTCCCTTCCCCTGTAG

Because this reminds you of me, dear so.

The worms crawl in and the worms crawl out

The ones that crawl in are lean and thin

The ones that crawl out are fat and stout

Your eyes fall in and your teeth fall out

Your brains come tumbling down your snout

Be merry my friends

Be merry

Pogues—Worms