Author/s:
Miller, AD;Good, RT;Coleman, RA;Lancaster, ML;Weeks, AR

Title:
Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and de novo genome assembly for the critically endangered orange-bellied parrot, Neophema chrysogaster

Date:
2013-01-01

Citation:
Miller, A. D., Good, R. T., Coleman, R. A., Lancaster, M. L. & Weeks, A. R. (2013). Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and de novo genome assembly for the critically endangered orange-bellied parrot, Neophema chrysogaster. Molecular Biology Reports, 40 (1), pp.35-42. https://doi.org/10.1007/s11033-012-1950-z.

Persistent Link:
https://hdl.handle.net/11343/282905

Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and *de novo* genome assembly for the critically endangered orange-bellied parrot, *Neophema chrysogaster*

Adam D. Miller[1, 2], Robert T. Good[3], Rhys A. Coleman [2,4], Melanie L. Lancaster[5] Andrew R. Weeks[1, 3]

[1]***cesar**, 293 Royal Parade, Parkville 3052, Australia*
[2]*Department of Zoology, The University of Melbourne, Parkville, Victoria, 3010 Australia*
[3]*Department of Genetics, The University of Melbourne, Parkville, Victoria, 3010 Australia*
[4]*Melbourne Water Corporation, 100 Wellington Parade, East Melbourne, Victoria 3001 Australia*
[5]*Healesville Sanctuary, Badger Creek Road, Healesville, Victoria 3777, Australia*

*Corresponding author: Adam D. Miller (amiller@cesaraustralia.com)

Keywords: *Neophema chrysogaster*, microsatellite library development, mitochondrial genome sequence, Next Generation DNA sequencing, *de novo* genome assembly

Abstract

A suite of polymorphic microsatellite markers and the complete mitochondrial genome sequence was developed by next generation sequencing (NGS) for the critically endangered orange-bellied parrot, *Neophema chrysogaster*. A total of 14 polymorphic loci were identified and characterized using DNA extractions representing 40 individuals from Melaleuca, Tasmania, sampled in 2002. We observed moderate genetic variation across most loci (mean number of alleles per locus = 2.79; mean expected heterozygosity = 0.53) with no evidence of individual loci deviating significantly from Hardy-Weinberg equilibrium. Marker independence was confirmed with tests for linkage disequilibrium, and analyses indicated no evidence of null alleles across loci. *De novo* and reference-based genome assemblies performed using MIRA were used to assemble the *N. chrysogaster* mitochondrial genome sequence with mean coverage of 116-fold (range 89 to 142-fold). The mitochondrial genome consists of 18,034 base pairs, and a typical metazoan mitochondrial gene content consisting of 13 protein-coding genes, 2 ribosomal subunit genes, 22 transfer RNAs, and a single large non-coding region (control region). The arrangement of mitochondrial genes is also typical of Avian taxa. The annotation of the mitochondrial genome and the characterization of 14 microsatellite markers provide a valuable resource for future genetic monitoring of wild and captive *N. chrysogaster* populations. As found previously, NGS provides a rapid, low cost and reliable method for polymorphic nuclear genetic marker development and determining complete mitochondrial genome sequences when only a fraction of a genome is sequenced.

Introduction

Next Generation Sequencing (NGS) has revolutionised the field of molecular biology through the rapid and cost-effective collection of large amounts of genomic data. While this technology has been applied widely across a variety of research disciplines [31], its utility has remained limited in the field of conservation genetics. NGS technologies provide an effective platform for the development of genetic markers that can be used to provide insight into population processes and the evolutionary history of species. This information is critical when devising optimal conservation strategies and is therefore being increasingly used to guide management decisions. More recently NGS has been used for the rapid and cost effective isolation of nuclear microsatellite markers, where it has been shown that random sequencing of a small fraction of a genome can result in a high density of potential microsatellite loci [10]. Polymorphic microsatellite loci are used widely in population genetics and are routinely used in the field of conservation genetics to identify individuals [21], determine relatedness between individuals [13], estimate gene flow and genetic structure between populations [5], determine genetic diversity estimates within populations [20] and also estimate effective [11] and census [32] population size.

Even though recent studies have demonstrated the utility of NGS for isolating microsatellite loci for species [18, 19], these studies rarely explore or utilize the bulk of NGS data. For example, a 454 sequencing analysis using only 1/8th of a 70 x 75 mm Pico Titer Plate will typically generate over 20 Mb of sequence data, yet microsatellite containing contigs often amount to less than 3% of total sequence reads (Miller & Weeks, unpubl. data). The remaining sequence data generally overlooked contains a high density of contigs of both nuclear and mitochondrial origin, many of which are potentially valuable genetic markers for systematic research. Mitochondrial DNA is particularly useful and is applied widely to explore patterns of intra- and interspecific genetic variation [5, 17]. By exploiting certain tissue types (e.g. blood or muscle), total genomic DNA extractions can contain high concentrations of mtDNA, which may then be overrepresented in NGS analyses [7]. Such an approach may allow entire mtDNA genome sequences to be generated when only a small portion of the genome is sequenced and at a fraction of the cost of traditional approaches. Here we test this method by extracting DNA from muscle tissue of the critically endangered orange-bellied parrot, *Neophema chrysogaster*, and undertaking a modest (1/8th of a 70 x 75 mm picoTitre Plate) 454 NGS analysis to isolate microsatellite loci and mtDNA sequences.

Neophema chrysogaster is endemic to south-eastern Australia, with a former range on the mainland extending from Adelaide (South Australia) through coastal Victoria and as far north as Sydney (New South Wales), and Tasmania extending along the west and south coasts, east to Bruny Island [6]. Since the 1920s, the species has suffered a steady decline in the wild, with major threats including the degradation and loss of habitat, and introduced predators and competitors [6], and the wild population is currently estimated to be less than 50 individuals (R. Pritchard, pers. comm.). A captive breeding program was initiated in the late 1980's, but to date this program has largely been unsuccessful in returning birds back to the wild. The orange-bellied parrot is protected by both State and Commonwealth legislation throughout its range, including a listing as 'Endangered' under the *Environment Protection and Biodiversity Conservation* (EPBC) *Act 1999*. The International Union for Conservation of Nature and Natural Resources (IUCN) lists the orange-bellied parrot as 'Critically Endangered'.

Microsatellite loci were developed primarily to help inform and monitor conservation efforts both in the captive breeding colony and the remaining wild populations. Similarly, we aimed to target mitochondrial DNA in our NGS approach so that future studies could use this resource to gain a better understanding of the evolutionary history of *Neophema chrysogaster*.

Material and methods

Next-generation sequencing

The 454 next generation sequencing platform was used to identify microsatellite and mitochondrial markers for *N. chrysogaster*. Approximately 10 µg of genomic DNA was extracted from muscle tissue from a single specimen using a QIAGEN DNA Easy kit (Qiagen). DNA was subsequently processed by the Australian Genome Research Facility (AGRF) where it was nebulized, ligated with 454 sequencing primers and tagged with a unique oligo sequence allowing sequences to be separated from pooled species DNA sequences using post-run bioinformatic tools. The DNA sample was analyzed using high throughput DNA sequencing on 1/8th of a 70 x 75 mm Pico Titer Plate using the Roche GS FLX (454) system [15].

Microsatellite isolation and characterisation

Unique sequence contigs possessing microsatellite motifs were identified using the software QDD version 2 [16]. Primer3 [29] was used to design optimal primer sets for each unique contig where possible. A selection of 55 contigs including di-, tri-, and tetra-nucleotide repeats, were used for subsequent analysis. Loci were screened for polymorphism using template DNA from eight individuals, representing three temporally spaced samples (1992, 2002, and 2005) from the wild population at Melaleuca, Tasmania. Loci were pooled into groups of four, labeled with unique fluorophores (FAM, NED, VIC, PET) and co-amplified by multiplex PCR using a Qiagen multiplex kit (Qiagen) and an Eppendorf Mastercycler *S* gradient PCR machine following the protocol described by Blacket et al. [3]. Genotyping was subsequently performed using an Applied Biosystems 3730 capillary analyzer (http://www.agrf.org.au) and product lengths were scored manually and assessed for polymorphisms using GeneMapper version 4.0 (Applied Biosystems).

Polymorphic loci were selected, pooled into two groups for multiplexing based on observed locus specific allele size ranges, and further characterized using DNA from 40 individuals sampled in 2002 from the Melaleuca wild population. Microsatellite profiles were again examined using GeneMapper version 4.0 and alleles scored manually. The *Excel Microsatellite Toolkit* [24] was then used to estimate expected (H_E) and observed (H_O) heterozygosities and number of alleles (N_A), while examination of conformation to Hardy-Weinberg equilibrium (HWE), the inbreeding coefficient (F_{IS}) and linkage disequilibrium estimates between all pairs of loci was conducted using GENEPOP version 4 [26]. Where necessary, significance values were adjusted for multiple comparisons using Bonferroni corrections [28]. Finally, all loci were assessed using MICRO-CHECKER to check for null alleles and scoring errors [34]. The frequency of null alleles per locus was obtained using the 'Brookfield 1' formula, as evidence of null homozygotes across loci was not observed [2].

Mitochondrial assembly and annotation

Sequence reads in SFF format were edited by trimming 454 primer tags using the Roche software. Sequence assembly of genomic sequence contigs was achieved by '*de novo* assembly' using the assembly software MIRA and default 454 parameters [4]. Annotation of the mitochondrial genome was determined using the DOGMA online software using the default parameters including a 5x parallel BLAST search option (http://dogma.ccbb.utexas.edu). DOGMA estimated gene positions, codon usages, transcriptional orientations, and where relevant, secondary structures. All alignments were confirmed by visual inspection with reference genome sequences. Genome annotation was exported to SEQUIN and submitted to Genbank (accession number JX133087). The software OGDRAW [14] was used to provide a visual depiction of the *N. chrysogaster* mitochondrial DNA gene content and orientation.

Results and discussion

Next-generation sequencing and de novo genome assembly

A total of 73,522 sequence reads covering 24.7 Mb of the *N. chrysogaster* genome was obtained by NGS. Previous studies indicate that these figures are not excessive as these are commonly achieved by NGS analyses using only 1/16th of a 70 x 75 mm picoTitre Plate [18, 19]. Nonetheless this data represents ~ 2% of the ~1.5 Gb parrot genome [1]. MIRA assemblies indicate that approximately 6.5% of the total sequence reads are of mitochondrial origin (4,765 reads). *De novo* assembly of mtDNA sequence contigs revealed complete genome coverage with a mean coverage of 116-fold (range 89-142).

Microsatellite isolation and characterization

A total of 1130 unique sequence contigs possessing microsatellite motifs were identified by QDD analysis, of which 883 contigs were found to possess optimal priming sites. Initially, 55 contigs were screened for polymorphism, with 39 containing di-nucleotide repeats, 12 containing tri-nucleotide repeats, and 4 containing tetra-nucleotide repeat motifs. The screening analysis found 14 loci to be polymorphic, 26 were monomorphic and 15 failed to amplify.

The majority of the 14 polymorphic loci were characterized by low to moderate genetic variation, with an average of 2.79 alleles per locus (range = 2 – 8 alleles) and heterozygosity estimates ranging between 0.06 and 0.74 (mean = 0.53). Linkage disequilibrium analyses confirmed maker independence (indicating no evidence of significant linkage between loci), while MICRO-CHECKER analyses revealed no evidence of null alleles or scoring issues across loci. All loci were found to conform with Hardy-Weinberg expectations and estimates of F_{IS} indicate no significant evidence of heterozygote excess or deficit. HWE and F_{IS} estimates for marker OBP55 are high, however, following Bonferroni corrections these were not significant.

Mitochondrial Genome of N. chrysogaster

Genome composition

The mitochondrial genome of *N. chrysogaster* is a circular molecule 18,034 bp in length, and characterized by a typical metazoan gene composition; 13 protein-coding genes, 2 ribosomal subunit genes (rRNA), and 22 transfer RNA genes (trn) (Figure 1; Table 2). The gene arrangement, including respective transcriptional polarities of genes, is typical of avian species and identical to those taxa described in Table 3. Five gene pairs were found overlapping by up to 6 bp (Table 2), a characteristic that has been reported for other animal mtDNAs including birds [35]. The majority-strand (α) encodes 28 genes, while the minority-strand (β) encodes 9 genes (Table 2). The nucleotide composition of the α-strand is 5,498 adenine (30.5%), 6,014 cytosine (33.3%), 2,546 guanine (14.1%), and 3,985 thymine (22.1%). While A-T biases of higher magnitude are commonly observed in other taxonomic groups such as arthropods and nematodes, more modest biases are common in birds, mammals and fish [30]. Bias to cytosine on the α-strand is a common feature of metazoan mtDNAs and is a feature that appears associated with the duration of single-stranded state of ‘heavy-stranded’ genes during mtDNA replication [27, 30].

A total of 2,662 noncoding nucleotides are evident in the genome, with 158 bp at 24 intergenic regions and a large 2,504 bp noncoding region (Table 2). The large noncoding region found represents the putative control region on the basis of its relative position between the trnQ and trnF which is typical of birds, and sequence characteristics (A+T-rich, noncoding). The *N. chrysogaster* putative control region is notably larger than those reported for species given in Table 3, however, control region length variations are common among avian species and other metazoan groups [35]. Gene lengths and A+T base compositions of the *N. chrysogaster* *α*-strand, protein-coding, rRNA, and trn genes, as well as the putative control region, are displayed in Tables 3.

Protein-coding genes

All protein-coding genes except for ND6 are encoded by the *α*-strand (Table 2), with overlapping nucleotides observed at the ATP6 and 8, and NAD4 and NAD4L gene boundaries (Table 2). Overlaps at these particular gene boundaries are a common feature of metazoan mitochondrial genomes [35], and have been validated by surveys of bicistronic transcripts and protein characteristics [9, 23]. Translation initiation and termination codons of the *N. chrysogaster* 13 protein-coding genes are summarized in Table 2. The standard methionine (ATN) initiation codon was inferred for 12 of the 13 genes while the ND5 gene appears to use a valine (GTG) codon, a nonstandard codon used in other metazoans including birds [8, 35]. Open reading frames are terminated with the typical TAA and TAG codons for all genes except for COIII and NAD4. We suggest that these genes are characterized by truncated termination codons (T) with the production of the TAA termini being created by post-transcriptional polyadenylation [23, 33]. This is a common feature reported for other metazoan mt genomes [12,17, 22].

Ribosomal and transfer RNA genes

The rRNA gene boundaries were estimated by BLAST sequence alignments implemented in DOGMA, with a high degree of conservatism at the beginning and end of the respective genes across avian taxa. Both ribosomal subunit genes are encoded by the α-strand with the *rrnS* (12S) gene separates trn F and trnV, and the *rrnL* (16S) gene separating the trnV, trnL$^{(uac)}$. The genomic position and transcriptional polarity of the rRNA genes is typical of avian species (Table 3).

A total of 22 trn genes corresponding with the standard set of metazoan genes were identified on the basis of their respective anticodons and secondary structures (Table 2). Gene lengths and anticodon sequences are largely congruent with those described for other avian species described in Table 3. All genes can be folded into the canonical cloverleaf structure except for trnS$^{(gcu)}$ and trnK which lack the DHU arm, instead replaced with unpaired loops 8 and 13 bases in length, respectively. Replacements loops are commonly observed in metazoan trnS genes [35]

Conclusion

The NGS approach using the 454 sequencing platform was successful in isolating 1130 microsatellite containing contigs for *N. chrysogaster* from a total of 73,522 sequence contigs that covered approximately 24.7 Mb of the genome. While birds are thought to have inherently low numbers of microsatellite loci [25], we successfully developed 14 polymorphic microsatellite markers that will be a valuable resource for devising effective conservation strategies for the species. These markers can be used to determine changes in genetic variation, relatedness, inbreeding, gene flow, genetic structure, effective population size and past population processes in both the wild and captive populations. They should also prove integral in guiding captive breeding programs, determining success of reintroductions, and assigning parentage in the wild. We genotyped all 14 loci from blood samples collected from 40 wild birds at Melaleuca in 2002, showing moderate to low levels of genetic variation as measured by estimates of heterozygosity and allelic richness. In 2004 the estimated population size of *N. chrysogaster* in the wild was thought to be less than 150 birds [6]. However, substantial declines have occurred since then and it is now estimated that less than 50 *N. chrysogaster* currently persist in the wild (R. Pritchard, pers. comm.). This highlights the importance of ongoing genetic monitoring of both wild and captive populations to inform on-going conservation efforts for this important and iconic species.

Interestingly, despite sequencing only a fraction of the nuclear genome (approx 2%), we were able to obtain an average coverage of 116-fold of the *N. chrysogaster* complete mitochondrial genome sequence. Extracting DNA from muscle tissue, which is inherently rich in mitochondria [7], likely resulted in an overrepresentation of sequence contigs of mitochondrial origin in the NGS analysis. We have demonstrated that by targeting specific tissues, the NGS analysis is a rapid and cost effective method for not only developing nuclear microsatellite markers, but also sequencing entire mitochondrial genomes. Combined these genetic markers are an extremely valuable resource for investigating the population genetics and evolutionary histories of endangered species, that in turn provides a framework for establishing effective conservation strategies

Acknowledgements

We thank Healesville Sanctuary for providing the *N. chrysogaster* specimen used to extract DNA for the NGS approach. Neil Murray is thanked for providing *N. chrysogaster* samples used for the characterization of microsatellite loci.

References

1. Andrews CB and Gregory TR (2009) Genome size is inversely correlated with relative brain size in parrots and cockatoos. Genome 52:261-267

2. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5:453-455

3. Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of 1 PCR fragments – an efficient and cost effective approach to genotyping by fluorescence. Mol Ecol Notes doi: 10.1111/j.1755-0998.2011.03104.x

4. Chevreux B, Wetter T, Suhai S (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99, pp. 45-56.

5. Coleman RA, Pettigrove V, Raadik TA, Hoffmann AA, Miller AD, Carew ME (2010) Microsatellite markers and mtDNA data indicate two distinct groups in dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae), a threatened freshwater fish from south eastern Australia. Conservation Genetics 11: 1911-1928

6. Commonwealth of Australia (2005) Orang-bellied parrot recovery plan. Commonwealth of Australia, Canberra.

7. Dalziel AC, Morre SE, Moyes CD (2005) Mitochondrial enzyme content in the muscles of high-performance fish: evolution and variation among fiber types. American Journal of Physiology Regulatory, Integrative and Comparative Physiology 228:R163-R172

8. Desjardins P, Morais R (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. Journal of Molecular Evolution 32(2):153-61

9. Fearnley IM, Walker JE (1986) Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO 5:2003–2008

10. Gardner MG, Fitch AJ, Bertozzi, T, Lowe AJ (2011) Rise of the machines – recommendations for ecologists when using next generation sequencing for microsatellite development. Molecular Ecology Resources 11(6):1093-1101

11. Gomez-Uchida D, Banks MA (2006) Estimation of Effective Population Size for the Long-Lived Darkblotched Rockfish *Sebastes crameri*. Journal of Heredity 97(6):603-606

12. Ki JS, Hwang DS, Park TJ, Han SH, Lee JS (2010) A comparative analysis of the complete mitochondrial genome of the Eurasian otter *Lutra lutra* (Carnivora; Mustelidae). Molecular Biology Reports 37(4):1943-55

13. Larson S, Christiansen J, Griffing D, Ashe J, Lowry D, Andrews K (2011) Relatedness and polyandry of sixgill sharks, *Hexanchus griseus*, in an urban estuary. Conservation Genetics 12(3):679-690

14. Lohse M, Drechsel O, Bock R (2007) Organellar Genome DRAW (OGDRAW) - a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics 52:267-274

15. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

16. Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26(3):403-404

17. Miller AD, Austin CM (2006) The complete mitochondrial genome of the mantid shrimp *Harpiosquilla harpax*, and a phylogenetic investigation of the Decapoda using mitochondrial sequences. Molecular Phylogenetics and Evolution 38:565-574

18. Miller AD**,** Versace, V.L., Matthews, T., Bowie, K.C. (2012) The development of 20 microsatellite loci for the Australian marine mollusk, *Donax deltoides*, through next generation DNA sequencing. Conservation Genetics Resources 4(2):257-260

19. Miller, A.D., van Rooyen, A., Ayres, R.M., Raadik, T.A., Fairbrother, P., Weeks, A.R. (2012) The development of 24 polymorphic microsatellite loci for the endangered barred galaxias, *Galaxias fuscus*, through next generation DNA sequencing. Conservation Genetics Resources DOI 10.1007/s12686-012-9605-x.

20. Mitrovski P, Heinze DA, Broome L, Hoffmann AA, Weeks AR (2007) High levels of variation despite genetic fragmentation in populations of the endangered mountain pygmy-possum, *Burramys parvus*, in alpine Australia. Molecular Ecology 16:75-87

21. Mondol S, Navya R, Athreya V, Sunagar1 K, Selvaraj VM, Ramakrishnan U (2009) A panel of microsatellites to individually identify leopards and its application to leopard monitoring in human dominated landscapes. BMC Genetics 10:79

22. Oh DJ, Kim JY, Lee JA, Yoon WJ, Park SY, Jung YH (2007) Complete mitochondrial genome of the rabbitfish *Siganus fuscescens* (Perciformes, Siganidae). DNA Sequence 18(4):295-301

23.Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474.

24. Park SDE (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. PhD Thesis, University of Dublin

25. Primmer CG, Raudsepp T, Chowdhary BP, Moller AP, Ellegren H (1997). Low frequency of microsatellites in the avian genome. Genome Research 7:471-482

26. Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280-1283

27. Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Molecular Biology and Evolution 15:957–966

28. Rice WR (1989) Analyzing tables of statistical tests. Evolution 43(1):223–225

29. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

30. Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238:195-209

31. Schuster SC (2008) Next-generation sequencing transforms today's biology. Nature Methods 5:16-18

32. Skaug HJ (2001) Allele-sharing methods for estimation of population size. Biometrics 57:750–756

33. Temperley RJ, Wydro M, Lightowlers RN, Chrzanowska- Lightowlers ZM (2010) Human mitochondrial mRNAs—like members of all families, similar but different.

Biochimica et Biophysica Acta 1797(6-7):1081–1085

34. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4(3):535–538

35. Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. International Review of Cytology 141:173–216

Table 1. Primer sequences and characteristics of 24 microsatellite loci isolated from *Neophema chrysogaster*.

Locus	Primer sequence (5'-3')	Repeat motif	*n*	N_A	Size range (bp)	H_O/H_E	HWE	F_{IS}	Frequency of null alleles	Genbank Accession number
OBP1	CTGGGATGAGGCAGACAAC TCTCTCAGCATTCCAGTGTG	(AC)^9	40	2	193-195	0.32/0.37	0.41	0.13	0.03	
OBP2	ACGTCACGGAGCCTTAGC CTCCTTAAGCTGCATGGTGG	(GT)^10	40	2	270-272	0.32/0.38	0.57	0.17	0.04	
OBP5	AGTTTACTGAAAGAACTACAGAGC TGCGCATGAAGTCAGCAAG	(AT)^9	40	2	255-257	0.30/0.26	1.00	-0.15	-0.03	
OBP9	CTGTGCACCAGAAAGAGATACC AGTGAAGTCTGGTTTGTGTCAG	(AC)^11	40	2	228-234	0.06/0.06	1.00	-0.02	0.00	
OBP10	GCAGAAAATAATGCTGTGCCTG ATGCATCCAGTGCCTGTTG	(AC)^8	40	2	170-172	0.34/0.42	0.26	0.20	0.06	
OBP13B	TGTGCTTCAGATTGCTCTGG TGGCTCTGGACTTACCAAGG	$(AC)^{11}$	40	4	107 - 109	0.65/0.56	0.79	-0.16	-0.06	
OBP16	CATTCTAAGCCCAGCCAGC ACTGCTACAATGGTCCTGC	(AC)^11	40	3	141-155	0.54/0.53	0.38	-0.02	-0.01	
OBP24	ACCCATGATTTCAGGCTTGC TATTGCTGCCGCCGTTTAC	(AC)^12	40	2	200-204	0.56/0.47	0.41	-0.20	-0.07	
OBP25	ACCCACCCACCAAACATAC GCACCCGTTACTTGGTCTTG	(AC)^9	40	2	232-238	0.48/0.46	1.00	-0.04	-0.02	
OBP35	CTCAGCAGGAGGGAGGTTG CCTTCTATCCGTCCACGTATCC	(GGAT)^9	40	3	200-208	0.67/0.66	0.73	-0.02	-0.01	
OBP36	GCATGTGAGCACTTGTTGC	(ATCT)^11	40	8	266-300	0.64/0.74	0.11	0.14	0.05	

	GCTCCCAGGATTCCCTTCC								
OBP38C	CCTGGCCAAATTGGATAACT	(AGG)^8	40	2	313-315	0.36/0.44	0.62	0.18	0.05
	ACCTCCTCCTCTCCCAACAT								
OBP54	ACAAGCTTAGCATACGCAGG	(AC)^8	40	2	206-208	0.43/0.50	0.66	0.15	0.04
	ATGGCAGGGGTTGGAACTG								
OBP55	CCACAACCTTTGCTCCCTC	(AC)^8	40	3	214-218	0.40/0.53	0.01	0.26	0.09
	TGCCAGCATGTTTAAAGGGG								

Number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities, Hardy-Weinberg equilibrium P values (HWE) and inbreeding coefficient (F_{IS}). H_O, H_E, HWE, F_{IS} and null alleles are estimated from a single population of 40 individuals.

Table 2. Mitochondrial gene profile of *Neophema chrysogaster*

Feature	Position number [a]		Size	Codon		Intergenic
	From	To	(bp)	Start	Stop	Nucleotides [b]
CO1	1	1545	1545	ATG	AGG	NA
trnS $^{Ser(uga)}$	(1540	1616)	77			-5
trnD $^{Asp(guc)}$	1620	1688	69			3
CO2	1691	2371	681	ATG	TAA	2
trnK $^{Lys(cuu)}$	2376	2444	69			4
ATP8	2446	2610	165	ATG	TAA	1
ATP6	2604	3284	681	ATG	TAA	-6
CO3	3287	4069	783	ATG	T*	2
trnG $^{Gly(ucc)}$	4071	4139	69			1
NAD3	4140	4487	348	ATA	TAA	0
trnR $^{Arg(ucg)}$	4490	4558	69			2
NAD4L	4560	4853	294	ATG	TAA	1
NAD4	4850	6241	1392	ATG	T*	-3
trnH $^{His(gug)}$	6243	6311	69			1
trnS $^{Ser(gcu)}$	6314	6376	63			2
trnL $^{Leu(uag)}$	6377	6447	71			0
NAD5	6448	8259	1812	GTG	TAA	0
CYTB	8274	9410	1137	ATG	TAA	14
trnT $^{Thr(ugu)}$	9413	9481	69			2
trnP $^{Pro(ugg)}$	(9490	9560)	71			8
NAD6	(9567	10082)	516	ATG	TAG	6
trnE $^{Glu(uuc)}$	(10085	10154)	70			2
Control region	10155	12658	2504			0
trnF $^{Phe(gaa)}$	12659	12723	65			0
rrnS (12S)	12723	13681	959			-1
trnV $^{Val(uac)}$	13701	13770	70			19
rrnL (16S)	13824	15346	1523			53
trnL $^{Leu(uaa)}$	15362	15435	74			15
NAD1	15442	16419	978	ATG	AGG	6
trnI $^{Ile(gau)}$	16421	16491	71			1
trnQ $^{Gln(uug)}$	(16500	16570)	71			8
trnM $^{Met(cau)}$	16570	16638	69			-1
NAD2	16639	17676	1038	ATG	TAA	0
trnW $^{Trp(uca)}$	17679	17749	71			2
trnA $^{Ala(ugc)}$	(17751	17819)	69			1
trnN $^{Asn(guu)}$	(17821	17895)	75			1
trnC $^{Cys(gca)}$	(17898	17964)	67			2
trnY $^{Tyr(gua)}$	(17964	18034)	71			-1

[a] Brackets denote β-strand transcriptional polarity

[b] Negative numbers indicating overlapping nucleotides between adjacent genes

* Incomplete termination codon likely extended via post-transcriptional adenylation

Table 3. Genomic composition of Avian mitochondrial DNA

		Genbank	*α* -strand		13 Protein-coding		2 rRNA genes		22 tRNA genes		Control region	
		accession	Length	A + T	Length	A +T	Length	A +T	Length	A +T	Length	A + T
Species	Family	number	(bp)	(%)	(bp)	(%)	(bp)	(%)	(bp)	(%)	(bp)	(%)
Neophema chrysogaster	Psittacidae	JX133087	18,043	52.6	11,403	51.8	2482	50.7	1,548	50.8	2504	55.1
Strigops habroptilus	Psittacidae	NC_005931	15,599*	56.8	11,402	57.0	2558	54.5	1,538	59.3	-	-
Melopsittacus undulatus	Psittacidae	NC_009134	18,193	55.3	11397	54.1	2549	51.6	1541	58.2	2685*	62.4
Brotogeris cyanoptera	Psittacidae	NC_015530	17,369	53.2	11386	52.6	2552	50.8	1536	55.9	1867	57.5
Ninox novaeseelandiae	Strigidae	AY309457	16,223*	52.8	11402	52.7	2557	51.1	1557	55.7	-	-
Anseranas semipalmata	Anseranatidae	AY309455	16,870	54.4	11399	54.6	2567	52.8	1547	58.0	1335	56.2
Gallus Gallus	Phasianidae	AP003580	16,788	53.9	11393	53.0	2601	53.3	1545	57.8	1232	60.0

* incomplete control region sequence

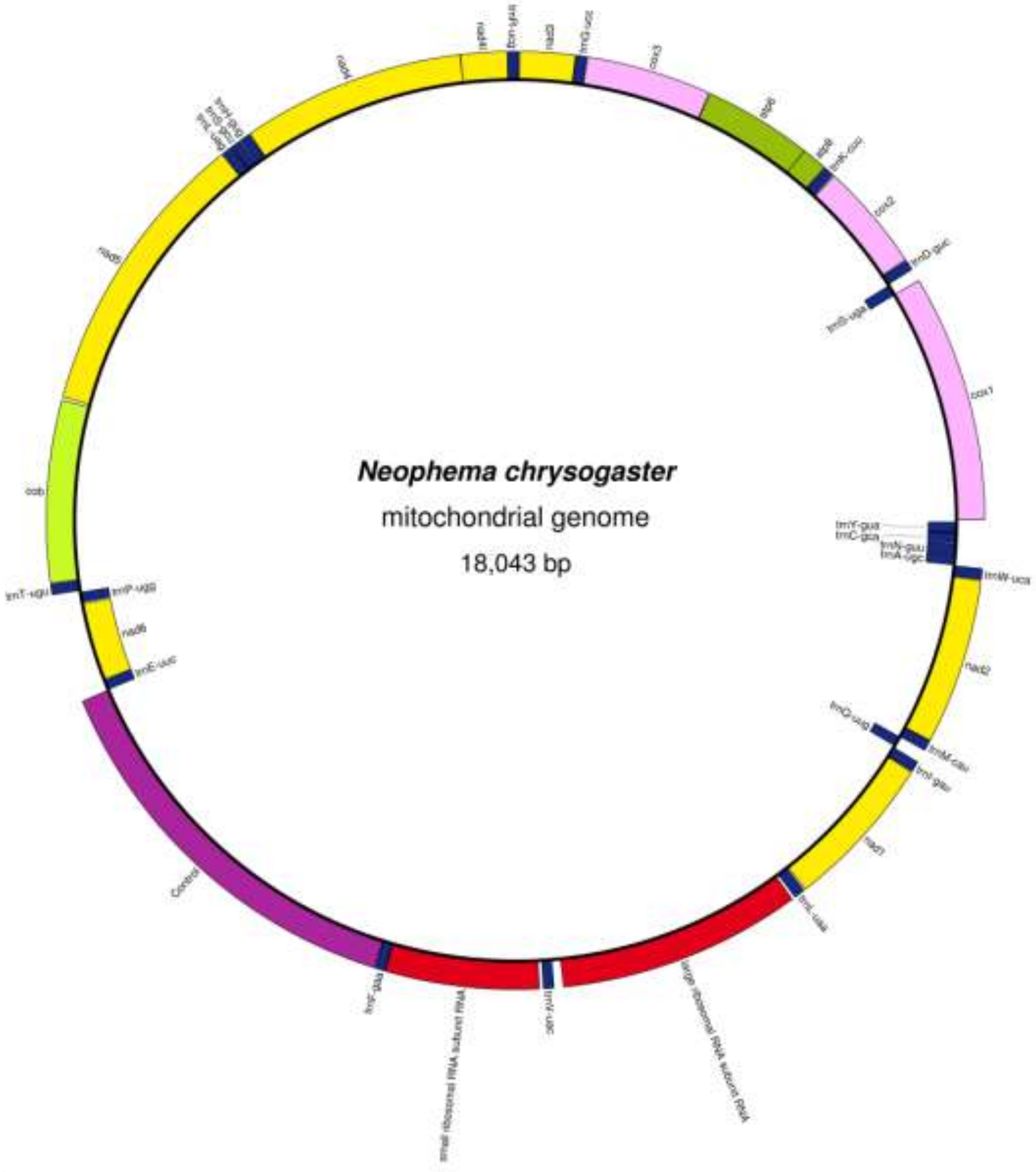

Figure 1. Gene map of the *Neophema chrysogaster* mitochondrial genome. COI-III indicates cytochrome c oxidase subunits 1–3; cyt b, cytochrome b; ATP6–8, ATPase subunits 6 and 8; ND1–6/4L, NADH dehydrogenase subunits 1–6/4L. Transfer RNA genes are designated by single-letter amino acid codes (Table 2.) Protein-coding and RNA genes on the outside circle are transcribed in a clockwise direction (α-strand), those encoded on the β-strand are shown on the inside circle of the molecule.