De novo sequencing, assembly and annotation of the *Agapornis roseicollis* genome to identify variants for the development of genetic screening tests

H Van der Zwan

orcid.org 0000-0001-5288-7609

Thesis submitted in fulfilment of the requirements for the degree *Doctor of Philosophy in Biochemistry* at the North-West University

Promoter: Dr R Van der Sluis

Co-promoter: Dr C Visser

Graduation July 2019

25621572

“There are not more than five musical notes, yet the combinations of these five give rise to more melodies than can ever be heard.

There are not more than five primary colours, yet in combination they produce more hues than can ever been seen.

There are not more than five cardinal tastes, yet combinations of them yield more flavours than can ever be tasted.”

— **Sun Tzu, The Art of War.**

There are not more than four nucleotides in the genome, yet the combinations of these create life.

Turquoise *A. fischeri* – Photo: Dirk Van den Abeele

ACKNOWLEDGEMENTS

I dedicate this thesis to my husband, Pieter, and our (very soon to be born) son, Emil. Pieter, your name means “rock” and to me that is what you have always been to me. Without you, I would never have taken this opportunity to pursue this degree. I also want to thank my Heavenly Father for giving me the right gene combinations to have an interest in genetics and the ability to study this field.

I would like to thank the following individuals and institutions, who all played a cardinal roll in this study:

Francois van der Westhuizen who was pivotal in the initial study design and obtaining funding.

My promotor, Rencia van der Sluis and co-promotor, Carina Visser for their guidance, support, expert knowledge and patience during the whole study.

Dirk Van den Abeele – everything I know of lovebirds I have learnt from you! I will be forever grateful for all the knowledge you have shared with me. Thank you so much for always helping me with whatever request I had and thank you also for hosting me in Belgium and introducing me to Kriek.

Thank you to BGI, Copenhagen and the Bird 10 K Consortium for the genome assembly and annotation. In particular a huge thank you to Erich Jarvis, Guojie Zhang, Tom Gilbert, Cai Li and Yuan Deng.

Thank you to Mario van Poucke for the DNA extractions when we required extra blood samples.

Maryke Schoonen and Bertie Seyffert, thank you so much for your help with all the bioinformatics analyses.

Thank you to all the lovebird breeders and owners who have sent in blood and feather samples. Your support is greatly appreciated.

Finally, the financial assistance from The Technology and Human Resources for Industry Programme (THRIP), The Technology Innovation Agency (TIA) and the National Research Foundation (NRF) has made this study possible.

SUMMARY

The genus *Agapornis* consists of nine small African parrots that are commonly called lovebirds. These birds are found in their natural habitat across Africa (eight species) and Madagascar (one species), but also house as domesticated pet birds across the globe. Eight of these species have been domesticated of which five are commonly found in breeding systems. Wild populations are placed under strain due to poaching and illegal export to sell birds to the pet market. Trade has subsequently been restricted in some countries due to declining numbers. Poaching, trapping and illegal export are however still a problem for some populations. The main selection criterion breeders use to select birds is plumage colour variations. There are 30 known colour variations found amongst these species, many of which can be combined. Very little research has been conducted on the molecular genetic mechanisms that control parrot plumage pigmentation. This group of birds have a unique pigment, psittacofulvin, that is only found amongst parrots. It is believed to be genetically controlled and not under dietary control. The inheritance pattern of these variations have been determined by breeders via test matings, and most are inherited as Mendelian traits. Despite the inheritance patterns being known, the genes and polymorphisms linked to these traits have not yet been identified. This has caused breeders to use pedigree data to predict the colour genotypes of an offspring with wildtype colouration. Notwithstanding this practice, there is no molecular parentage verification panel available for lovebirds. The avian parentage verification panels that are available were developed for an array of bird species and are all microsatellite marker based. Microsatellite markers are becoming redundant in animal breeding systems and replaced with Single Nucleotide Polymorphism (SNPs). One of the limitations of developing a parentage verification panel for this genus, is the lack of a reference genome from where SNPs could be identified. The *de novo* genome of *A. roseicollis* were subsequently sequenced, assembled and annotated for this purpose. Sequencing was performed at 100x coverage using the Illumina HiSeq 2000 platform. Three shotgun sequencing libraries of insert sizes 300 bp, 550 bp and 750 bp, respectively, as well as two long jumping distance paired-end libraries of sizes 3 kbp and 8kbp were constructed. The *de novo* assembly was performed using the SOAPdenovo v2.04 assembler and a *k*-mer length of 69 was applied. The genome was found to be 1.1 Gbp in size, with contig and scaffold N50 lengths of 5 45 bp and 108 514 bp, respectively, and the G/C content 43%. During the genome annotation phase 15 045 coding gene sequences and 999 non-coding gene sequences were identified. The genome assembly compared well with previously assembled avian genomes such as the budgerigar (*Melopsittacus undulates*), scarlet macaw (*Ara macao*) and Puerto Rican parrot (*Amazona vittata*) in terms of genome size, number of genes annotated and scaffold and contig N50 lengths. The number of eukaryotic core genes detected in the lovebird assembly outperformed

those identified in the budgerigar, Puerto Rican parrot and scarlet macaw assemblies, indicating that the assembly was accurate and complete.

The genomes of both of the parents of the reference genome individual were sequenced and the sequencing data used to identify SNPs throughout the genome that could be included in a parentage verification panel. Sequencing was performed at 30x coverage using the Illumina HiSeq 2000 platform. Two shotgun sequencing libraries of insert sizes 300 kb and 550 kb, respectively, were constructed. These reads were mapped against the reference genome of their chick and variants were discovered using the variant caller Genome Analysis Toolkit (GATK). Over 2 million raw variants were discovered for the mother while 1,60 million raw variants were discovered for the father. The parents' genotypes were combined to identify SNPs that were shared by the two birds. Unwanted variants such as insertions and deletions (indels) were discarded which resulted in a callset of 1,66 million SNPs. These SNPs were filtered based on parameters as recommended by GATK resulting in 103 287 SNPs that passed the criterion that were set. Two of the parameters applied were QUAL (a Phred-based prediction of a false positive variant) and QD (normalization of the QUAL score for sequencing depth). True variants are found in the QD range of 11.5 to 12.5 and a higher QUAL score indicate a true variant. Therefore, all SNPs within this QD range, subsequently ranked by their QUAL scores were included. One SNP per scaffold was selected from this set and the top 480 SNPs were included in the final parentage verification panel. A population of 960 lovebirds from seven different species were genotyped at these 480 SNPs using the QuantStudio 12 K Flex platform. These birds included the reference genome individual and its father. A panel of 262 SNPs were constructed where the father’s genotype amplified and were used as a reference. This panel was further reduced to include SNPs with minor allele frequencies (MAF) and observed heterozygosity (H_O) values greater than 0.1. This resulted in a panel of 195 SNPs. The third panel was filtered based on the same parameters but included SNPs with MAF and H_O values greater than 0.3 and amounted to 40 SNPs. The three panels were all assessed for their exclusion power in 43 lovebird families with known pedigrees. It was found that the 195-SNP panel was the panel with the greatest exclusion power applying the least number of SNPs and was proposed as the lovebird parentage verification panel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 STUDY MOTIVATION AND RATIONALE

Agapornis (or lovebirds) is a genus consisting of nine small parrot species housed worldwide as popular pet birds but also found in their natural, wild habitat across Africa and Madagascar (Van den Abeele, 2016; Forshaw, 1989). Lovebirds are bred across the globe with large breeding societies located in Europe, Asia, the United States of America and Africa. Unlike other animal breeding industries, there are limited genetic screening tests available for these birds. This is largely due to the absence of a reference genome for any of the nine *Agapornis* species from which Single Nucleotide Polymorphisms (SNPs) and genes of economic importance can be identified.

In lovebird aviculture the main selection criterion used by breeders is that of plumage colour variations. There are 30 known colour variations and many of these can be combined to create even more spectacular plumage colours (Van den Abeele, 2016). The mechanism of parrot colouration is known and involves a pigment called psittacofulvin as well as structural changes in the feather barbs (Mundy, 2006 b; Dyck, 1971 a and b). These mechanisms are believed to be genetically controlled and it was established by breeders that the variations are inherited as Mendelian traits. The genes and polymorphisms linked to the traits are yet to be discovered. Therefore, breeders make use of pedigree information to predict if a bird with wildtype coloration might be a heterozygote of a specific colour. Since no species-specific or genus-specific molecular pedigree verification tests are routinely available, buyers cannot dispute a seller's claims of either the bird's pedigree or possible colour genotype. Breeding societies also don't require breeders to verify parentage before a bird is sold, imported or exported. Fraudulent transactions often occur where a bird is sold (frequently at a premium) without scientific evidence of the pedigree or colour genotype. Close relatives are also regularly mated to ensure that recessive alleles are combined to produce desired colour variations in their offspring which could lead to a serious lack in genetic diversity. The development of a genus-specific SNP-based parentage verification test for lovebirds has been long overdue. One of the limitations has been the identification of SNPs.

Currently, available parrot parentage verification panels are all microsatellite based panels (Coetzer *et al.,* 2017; Jan & Fumagalli, 2016) and were developed for other parrot species. Microsatellite markers have been widely used to verify parentage in many domesticated animal species, but the use of these markers has generally become redundant and it has been replaced with Single Nucleotide Polymorphisms (SNPs). SNPs are bi-allelic and therefore more SNPs are needed in a parentage verification panel compared to microsatellite markers

(Olsen *et al.*, 2011). SNPs, on the other hand, amplify more consistently than microsatellite markers making data sharing between laboratories not always possible (Vignal *et al.*, 2002; Garvin *et al*., 2010). Amplification of microsatellite markers across species is not always successful. Consequently, this highlights the need of the development of a SNP-based, genus specific parentage verification tests for lovebirds.

Many of the wild populations of some of the lovebird species are under pressure due to, amongst others, illegal trade, poaching and trapping (Perrin, 2012). Even though trade has been banned in many African countries, birds are still poached from their nests and exported on the black market. The development of genetic screening tests could aid conservationists and custom officials to identify the identity, and in future, species of confiscated tissue samples, eggs or live birds that could in turn lead to prosecution of poachers.

The sequencing, assembly and annotation of the lovebird genome will set the foundation from which genetic screening tests such as parentage verification, species identification and plumage coloration genotyping can be developed. Although not all these topics can be addressed in the current study, the sequencing, assembly and annotation of the genome and the development of a genus-specific, SNP-based parentage identification panel for all domesticated lovebird species will be discussed.

1.2 AIMS AND SPECIFIC OBJECTIVES OF THIS STUDY

1.2.1 MAIN AIMS

a. Sequence, assemble and annotate the *A. roseicollis* genome.
b. Whole Genome Resequencing (WGR) of the parents of the reference genome individual to discover SNPs.
c. Genotyping of a SNP panel in a larger lovebird to determine if these SNPs amplify and are polymorphic across all species, and compile a SNP panel for parentage verification.

1.2.2 SPECIFIC OBJECTIVES

i. To sequence the whole genome of one *A. roseicollis* individual at 100x coverage, to assemble these sequenced reads into a draft genome and finally to annotate the draft genome.
ii. To sequence the parents of the reference genome individual at 30x coverage and to map the parents' reads to the reference genome. Thereafter, to utilize GATK to identify variants throughout the genome.

iii. To filter variants to compile a panel of 480 SNPs most suited to be included in a parentage verification panel.
iv. To genotype these 480 SNPs in a population of 960 lovebirds using the Thermo Fischer QuantStudio 12K Flex platform.
v. To determine the optimum number of SNPs to include in the parentage verification panel that will have the highest exclusion power.

1.3 STRUCTURE OF THIS THESIS

This thesis is presented in five chapters that include two peer-reviewed articles.

Chapter 2: Literature review.

A literature overview that includes a discussion of the different lovebird species in their wild habitat as well as in aviculture. The different plumage colour variations found in this genus as well as the mechanisms how these variations are expressed are discussed in depth. This review covers *de novo* genome sequencing, assembly and annotation with special focus on studies conducted on avian species and, lastly, SNP identification as well as criteria to compile a SNP-based parentage verification panel are also discussed.

A review article (Published Paper I) that focusses on the problems wild lovebird populations face as well as an in depth discussion on the plumage colour variations found amongst lovebirds was incorporated into Chapter 2.

This manuscript was featured as the lead article of the journal's issue.

- Published paper I: **Plumage colour variations in the *Agapornis* genus: a review.** **Henriëtte van der Zwan**, Carina Visser and Rencia van der Sluis

  Published in: Ostrich (2019) 90:1, 1-10.

The full published article is attached in Appendix A.

Chapter 3: *De novo* sequencing, assembly, annotation as well as assessing the quality of the genome of *A. roseicollis*.

This chapter motivates the methods used during the sequencing, assembly and annotation phase of this study by discussing the methods used in previous avian *de novo* genome sequencing studies. A peer-reviewed paper describing the sequencing, assembly and annotation of the genome of *A. roseicollis* and a comparison to other avian *de novo* genomes, is shown.

- Published Paper II: **Draft *de novo* genome sequence of *Agapornis roseicollis* for application in avian breeding.**
  Henriëtte van der Zwan, Francois van der Westhuizen, Carina Visser and Rencia van der Sluis.

  Published in: Animal Biotechnology (2018) 29:4, 241-246.

Chapter 4: **Development of a SNP-based parentage verification panel for lovebirds.**

This chapter discusses the pipeline followed using the variant caller Genome Analysis Toolkit (GATK) to discover SNPs by Whole Genome Resequencing (WGR) of the two parents of the reference genome individual in more detail. This was done by whole genome resequencing (WGR) of the parents of the reference genome chick and identifying SNPs from their reads using GATK. A panel of 480 SNPs were compiled from the identified SNPs and genotyped in a population of 960 birds from various species. A wider population was tested to establish if the SNPs amplified across all lovebird species and if they were polymorphic. A final panel of SNPs were constructed and assessed based on heterozygosity (H), minor allele frequencies (MAF) and exclusion power of the individual SNPs and as a panel.

- Submitted Manuscript I: Manuscript submitted to Animal Genetics (Manuscript ID: AnGen-19-03-0071):
  Development of a SNP-based parentage verification panel for lovebirds.
  Henriëtte van der Zwan, Maryke Schoonen, Carina Visser and Rencia van der Sluis.

Chapter 5: Summary and Conclusion.

This chapter includes the general discussion, recommendations and conclusions of the study as well as future projects that could develop from this study.

References

All references are given at the end of the thesis.

Materials and Methods

The materials and methods used in this study are described in the Materials and Methods sections of the published Paper I and Submitted Paper I. Additional materials and methods not discussed in these manuscripts are given in Chapter 3 and 4.

1.4 APPENDIXES

Appendix A: Published Paper I: **Plumage colour variations in the *Agapornis* genus: a review. Henriëtte van der Zwan**, Carina Visser and Rencia van der Sluis. Published in: Ostrich (2019) 90:1, 1-10.

Appendix B: An example of the consent forms used for Submitted Manuscript I.

Appendix C: List of conference proceedings, abstracts and scientific posters.

Appendix D: Materials and methods as well as results not shown in Paper II.

Appendix E: Supplementary files submitted as part of Submitted manuscript I (Development of a SNP-based parentage verification panel for lovebirds.)

1.5 AUTHOR CONTRIBUTIONS

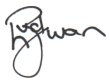
Paper I presented in Chapter 2: Henriëtte van der Zwan was involved in the review of the literature and manuscript writing of all sections. Carina Visser and Rencia van der Sluis were involved in the review of the literature, writing of the manuscript, revision and supervision.

Paper II presented in Chapter 3: Henriëtte van der Zwan was involved in the study design, sample collection and manuscript writing. Francois van der Westhuizen was involved in the supervision and study design. Carina Visser and Rencia van der Sluis were involved in the study design, manuscript writing, revision and supervision.

Submitted Manuscript I in Chapter 4: Henriëtte van der Zwan was involved in the design of the study, sample collection, pre-amplification of DNA samples, data analyses, SNP selection, assay design and manuscript writing. Carina Visser was involved in study design, manuscript writing and revision. Rencia van der Sluis was involved in the study design, assay design, manuscript writing and revision. Maryke Schoonen was involved in the bioinformatics pipelines during the GATK analysis.

All authors involved signed the declaration on this page:

As a co-author/researcher, I hereby approve and give consent that the above-mentioned articles and data can be used for the PhD thesis of Henriëtte van der Zwan. I declare that my role in the study, as indicated above, is a representation of my actual contribution.



Mrs. H. van der Zwan

Dr. R. van der Sluis

Dr. C. Visser

Prof. F.H. van der Westhuizen

Dr. M Schoonen

CHAPTER TWO: LITERATURE OVERVIEW

2.1 INTRODUCTION

There are 356 extant psittaciform, or parrot, species found globally (Forshaw, 1989) of which twenty are native to the African continent and Madagascar (Perrin, 2009). Amongst this order, the genus *Agapornis* (Selby, 1836) is found and consists of nine species, namely *A. roseicollis*, *A. taranta*, *A. canus*, *A. personatus*, *A. lilianae*, *A. swindernianus*, *A. fischeri*, *A. pullarius* and *A. nigrigenis* (Moreau, 1948; Dilger, 1960; Forshaw, 1989). These species are more commonly known as lovebirds and are distributed across Madagascar (*A. canus*) and Africa (the remaining eight species) (Dilger, 1960; Forshaw, 1989; Perrin, 2012). In addition to existing in their natural habitat, lovebirds are popular pets and eight of the nine species, *A. swindernianus* being the exception, are kept and bred as domesticated birds (Hayward, 1979; Silva & Kotlar, 1988; Forshaw, 1989; Van den Abeele, 2016).

Lovebirds have been kept as pets since the eighteenth century with reports of *A. pullarius* being bred in Britain in the 1880's (Farner *et al.*, 1982). Hayward (1979) states that lovebirds are popular in aviculture because they breed relatively easy in captivity, are hardy, active and have a range of plumage colours. They can also be bred in a small aviary making them ideal birds for a modern lifestyle. As a consequence of their popularity in aviculture, birds in their natural habitat are being threatened by poaching and trapping for export to pet markets (Warburton & Perrin, 2005 & 2006; Mzumara *et al.*, 2016 a & b). It is estimated that 123 species, or 34.6% of all parrot species, are listed as near-threatened to endangered (Forshaw, 1989; Pires, 2012). The International Union of Conservation of Nature (IUCN) has classified *A. nigrigenis* as vulnerable, *A. fischeri* and *A. lilianae* as near threatened and the other *Agapornis* species as "least concern". With the exception of *A. roseicollis*, all species are classified by The Convention of International Trade in Endangered Species of Wild Fauna and Flora (CITES) as Appendix II species indicating that trade is restricted to protect the species.

Despite their popularity, unlike other pet breeding systems such as dogs, where parentage (Qiu *et al.*, 2016) and coat colour heterozygosity (Schmutz & Berryere, 2007) are routinely tested and the results incorporated into the breeding system, there are only a handful of genetic screening tests available for parrots. The molecular screening tests most frequently used include gender determination (especially for birds where genders are similar in colour) (Morinha *et al.*, 2012) and pathogen testing to identify the beak and feather disease virus (BFDV) causing Psittacine beak and feather disease (PBFD) (Heath et al., 2004). Pedigree data is not verified using molecular genetic tests. There are parentage verification tests

available for avian species but none of these were developed for any *Agapornis* species. Plumage colour variations are the main economic factor breeders select for and there are at least 30 colour variations amongst domesticated lovebirds that are accepted by shows and auctions (Hayward, 1979, Van den Abeele, 2016). Rare colour variants are sold for up to 700 times as much as a wild type coloration bird of the same species in Europe (Personal communication: Mr. Dirk Van den Abeele) and there are reports of mark-ups of up to 1300 times in Asia (Diega, 2017). All of these colour variants are genetically inherited, but to date the only indication of the mode of inheritance of these traits are breeders' records. This highlights the need for the development of a parentage verification test that can routinely be applied in lovebird breeding systems. This will lower the number of fraudulent transactions and improve conservation efforts. In order to develop these tests, genetic markers and genes first need to be identified from the genome. The sequence, assembly and annotation of a reference genome will aid in the ease to identify these genomic elements.

2.2 AGAPORNIS SPECIES IN THEIR NATURAL ENVIRONMENT

A. roseicollis, *A. fischeri*, *A. personatus*, *A. lilianae* and *A. nigrigenis* together with the Monk Parakeets (*Myiopsitta monachus*) are the only parrot species to constructs their own nests (Eberhard, 1998; Warburton & Perrin, 2005; Ndithia *et al*., 2007). These five lovebird species carry nesting material of bark, leaves or grass in their rump feathers (*A. roseicollis*) or beaks (*A. fischeri*, *A. personatu*s, *A. lilianae* and *A. nigrigenis*) to build cup-shaped nests within cavities (Forshaw, 1989; Eberhard, 1998; Ndithia *et al.*, 2007; Warburton & Perrin, 2005). The other *Agapornis* species use tree holes as nests (*A. taranta* and *A. canus*) or cavities in arboreal ant or termite nests (*A. pullarius*) (Forshaw, 1989; Eberhard, 1998). The nesting behaviour of *A. swindernianus* is unknown.

A short description of each of the nine species as well as hybrid birds follows below. Photos of the eight domesticated *Agapornis* species, with their wildtype plumage coloration, are shown in Figure 2.1 a-h.

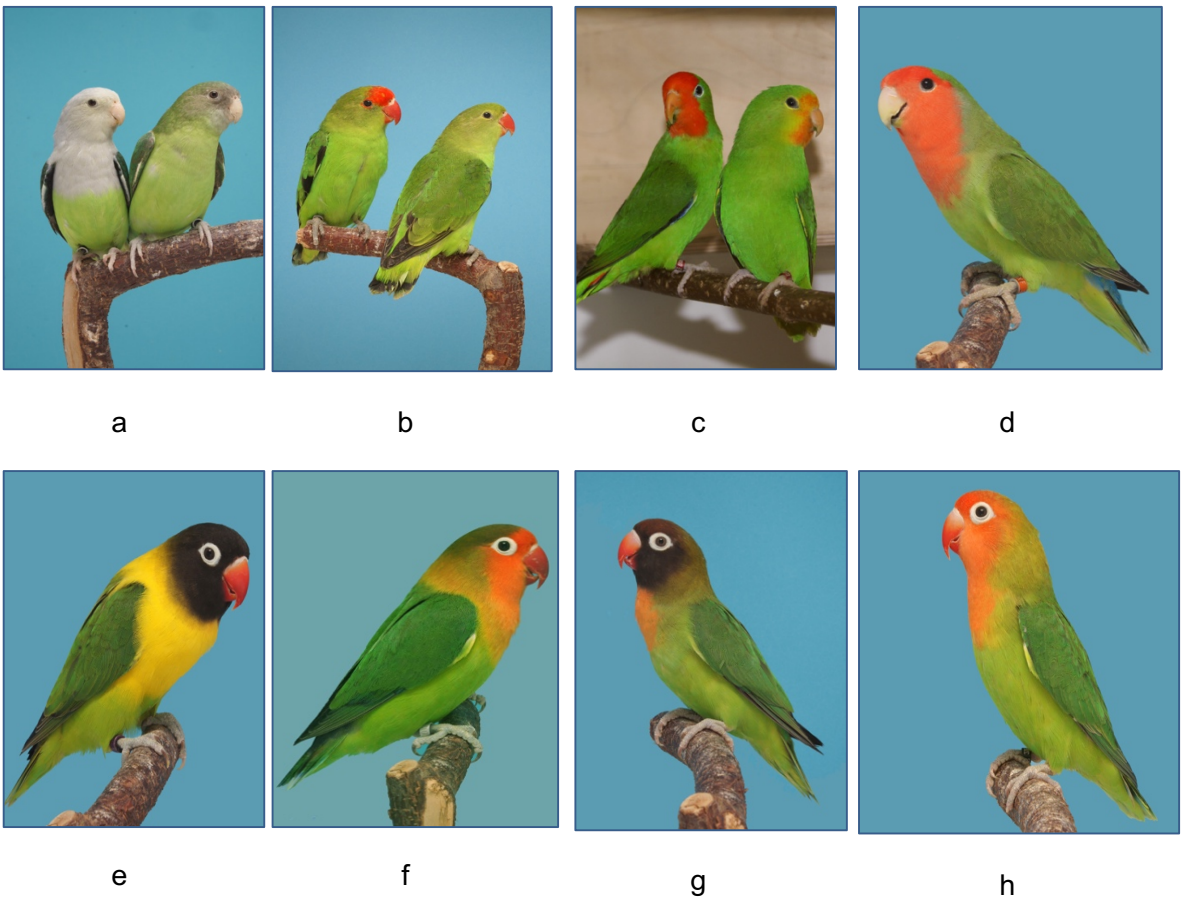


Figure 2.1: Eight domesticated *Agapornis* species with wildtype coloration. (Photos: Mr Dirk Van den Abeele). a. (*A. canus*), b (*A. taranta*) and c (*A. pullarius*) shows the three species from the sexually dimorphic group and displays the phenotypic plumage difference between the males (left) and females (right) for these species. Figure d illustrates that *A. roseicollis* is the species in the intermediate group as it displays traits of both groups. Males and females look identical but the white ring around the eye is absent. Figure e (*A. personatus*), f (*A. fischeri*), g (*A. nigrigenis*) and h (*A. lilianae*) shows the four species from the white eye ring group. Males and females are identical and the white skin around the eyes is clearly visible.

2.2.1 *AGAPORNIS CANUS*

A. canus is the only lovebird species inhabiting Madagascar (Farshaw, 1989) and is also called the grey-headed, lavender-headed or Madagascar lovebird. These birds are found amongst the Madagascar coastal areas and deciduous forests. Forshaw (1989) described this species as fairly tame, noisy and gregarious flock birds of up to 30 birds. At about 14 cm, *A. canus* is the smallest of the species and the male and females are dimorphic, as can be seen in Figure 2.1a. Perrin (2009) describes this species as a lowland habitat generalist that feeds largely on grass seeds. The male has a grey head, neck and chest while the rest of the body is green. The female is almost completely green with an ash-grey shine and both sexes have a white beak (Hayward, 1979; Silva & Kotar, 1988; Farshaw, 1989; Van den Abeele, 2016). Two

slightly differentiated sub-species are recognised (Farshaw, 1989; Van den Abeele, 2016) – *A. c. canus* ranging all over Madagascar except the south-western arid zone, and *A. c. ablectaneus* found in south-western Madagascar. The males of the latter have a deeper grey head, neck and chest and has a violet shine over their bodies (Van den Abeele, 2016). *A. canus* does not easily breed in captivity and no known colour variations exist for this species.

2.2.2 *AGAPORNIS TARANTA*

A. taranta is found in the higher laying areas of central and western Ethiopia, southern Eritrea and eastern Djibouti (Farshaw, 1989; Van den Abeele, 2016). At 17cm in height they are the largest of the lovebird species and males and females are dimorphic in colour but equal in size (Marez, 2003). The male has red plumage from the beak to the top of the skull and has a red ring around the eye which is absent in the female (Figure 2.1 b). The male's primary flight feathers are black with bluish-black wing edges and the tail has a black ring at the end, which is absent in the female. Both male and female are otherwise green with a red beak. Forshaw (1989) described them as resident with some seasonal movement dictated by food sources and up to 20 birds are commonly found in a flock. Tekalign & Bekele (2011) studied the status of *A. taranta* in Ethiopia and found many of the birds living in human habitats in the Bole Sub-City, probably migrating to better food sources. Their study also concluded that the numbers of this species might have dropped in the last couple of years, mainly due to depletion of their natural habitat. Commonly called the black-winged or Abyssinian lovebird, this species feed on tree fruits including figs and berries (Perrin, 2009). This species is not popular in aviculture, most possibly since they do not display a significant range of colour variations.

2.2.3 *AGAPORNIS PULLARIUS*

Since they don't breed as easily in captivity as other lovebird species, *A. pullarius*, or commonly called the red-faced or red-headed lovebird, is the least popular pet species of the eight domesticated species (Van den Abeele, 2016) and is often only obtainable at a very high price (Silva & Kotlar, 1988). One of the reasons leading to not being successful breeders in aviculture is that they breed in termite nests in their natural habitat and do not readily accept artificial nests (Silvia & Kotlar, 1988). It is suggested that *A. pullarius* was the first of the *Agapornis* species to be imported to Britain in the sixteenth century (Hayward, 1979). *A. pullarius* can be found across the African equator in 26 African countries, making it the most widespread of the lovebird species (Egwumah & Iboyi, 2017). Forshaw (1989) described them as resident birds found in flocks of less than 20 birds, but larger flocks are found amongst ripened crops. This species is shy and difficult to approach which could be another reason for its unpopularity as domesticated pets. Birds are approximately 15 cm tall and males and females are dimorphic (Van den Abeele, 2016). Males have an orange red forehead, crown,

cheeks and chest and the females have a pale orange red forehead, crown, cheeks and chest, as can be seen in Figure 2.1 c. The plumage colour on the bodies of both males and females are predominantly green. The males are black to dark ultra-marine blue from the wing bend onwards with a few sky blue feathers, but this is absent in the female. The primary flight feathers of the male are black and the rump is sky blue (Forshaw, 1989). Two poorly differentiated sub-species are found namely *A. p. pullarius* and *A. p. lugandae*. The latter only differs from the former by a paler blue lower back and rump with its distribution being limited to East-central Africa (Hayward, 1979; Forshaw, 1989). No plumage colour variations are known for either of these sub-species. Wild flocks feed on seeds of tall grasses that are often eaten when still green (Perrin, 2009).

2.2.4 *AGAPORNIS ROSEICOLLIS*

A. roseicollis's natural habitat is found in Angola, Namibia, Botswana and South Africa. Some feral populations are distributed over South Africa and in Phoenix and Tucson, Arizona in the United States of America (Forshaw, 1989; Symes, 2014). This species is 16 cm in height and males and females look identical. They are primarily green with a dark red mask that gradually becomes deep pink from the forehead to below the beak as can be seen in Figure 2.1 d. The rump and upper tail are a deep sky blue (Silva & Kotlar, 1988). Two poorly differentiated sub-species are recognised namely *A. r. roseicollis* and *A. r. catumbella* (Forshaw, 1989). Wild *A. roseicollis* flocks consist of less than 20 individuals but could reach a couple of hundred birds when close to water sources and at night as they roost communally. *A. roseicollis* is one of the most popular pet species and with twenty different plumage variations the peach-faced or rosy-faced lovebird is the species with the most known colour variations (Hayward, 1979; Forshaw, 1989; Van den Abeele, 2016). In their natural habitat, *A. roseicollis* build cup-shaped nests with cavities (Eberhard, 1998; Ndithia *et al.*, 2007) in *Acacia* species trees, telephone poles and human structures (Ndithia *et al.,* 2007). Perrin (2009) describe their diet as preferably seeds from grasses like *Anthephora schinzii*. However, many other grass seeds and fruits for example, seeds from *Albizia* (silk plants) and *Acacia* trees, fruits of *Ziziphus mucronata* (buffalo thorn), *Rhus villosa*, *Commiphora* spp. and *Ficus* spp. (figs) are also consumed.

2.2.5 *AGAPORNIS PERSONATUS*

The habitat of *A. personatus* lays in north-central Tanzania, north of Mount Meru south to Mbeya and Njombe districts with feral populations found at Tanga, Dar es Salaam, Nairobi, Mombasa and coastal areas (Forshaw, 1989; Van den Abeele, 2016). These birds are 15 cm tall, have a pitch black head and yellow chest and neck band, from which the common names "masked or black-masked lovebird" originated. It has a distinct white ring around its eye (one

of the four species belonging to the "white eye ring group") as shown in Figure 2.1e. The body is green with the wing coverts a little darker than the rest of the body. The wing bend has a yellow edge and the rump is mauve. Males and females have no distinguishable differences. They are one of the most popular species in aviculture as they breed easily in captivity and display six colour variations. It was also the first *Agapornis* species where the blue phenotype was recorded. Perrin (2009) list their diet as consisting of grass seeds. In the wild they are found in well-timbered grassland or bushlands preferring *Acacia*, *Commiphora* and *Adansonia* trees and feral populations thrive in gardens (Forshaw, 1989). They are known to live in smaller flocks of up to 20 birds but nest communally in large flocks.

2.2.6 *AGAPORNIS FISCHERI*

This species was named after Dr Fischer (commonly called Fischer's lovebird) who first described it in northern Tanzania (Van den Abeele, 2016). This sexually morph species stand at 15 cm in height with an orange red mask on the crown that becomes lighter below the red beak as shown in Figure 2.1 f. The Fisher's lovebird has a patch of clear white skin around its eye. Overall the body is green with blue at the end of the tail feathers. The blue rump differentiates it from *A. lilianae* (Forshaw, 1989). Mwangomo *et al.* (2007) described the habitat of this species as the Serengeti grassland region and their habitat is listed as North-central Tanzania from Kome and Ukerewe islands, southern Lake Victoria, Serengeti and Arusha National Parks south to Nzewga and Singida up to eastern Rwanda and Burundi with various feral populations (Forshaw, 1989; Symes, 2014). Wild birds feed of a diet containing mainly grass seeds and seeds of Acacias, fallen berries and figs (Perrin, 2009). *A. fischeri* is considered as one of the three most popular species in aviculture and display eighteen known plumage colour variations.

2.2.7 *AGAPORNIS NIGRIGENIS*

Inhabiting Zambia, the species name for *A. nigrigenis* is a combination of "niger" (black) and "genis" (cheeks) therefore the "black cheek" lovebird (Van den Abeele, 2016). Both males and females are 14 cm tall with a rusty brown brow and forehead which fade into dark brown as shown in Figure 2.1 g. The chin, throat and cheeks are black and the back of the head is olive green and this species is a member of the white eye ring group. The body is a dull grass green whereas the lower chest, abdomen, sides and cloaca are yellowish green. (Warburton & Perrin, 2005; Van den Abeele, 2016). This species is classified as vulnerable as wild population numbers are decreasing due to a number of factors including trade, water scarcity in their natural habitat and human threat since lovebirds feed off crops (Warburton & Perrin, 2005; Warburton & Perrin, 2006). Black-cheeked lovebirds are commonly found amongst mopane trees on flat alluvial soils of river valleys in small flocks of up to 15 birds (Forshaw,

1989). Their diets consist of annual grass and herb seeds and seeds from trees like *Albizia*, *Rhus* and *Combreumn* spp. *A. nigrigenis* is less popular in aviculture and only display a few colour variations.

2.2.8 *AGAPORNIS LILIANAE*

A. lilianae (also called Nyasa lovebird or Lilian's lovebird) is found in three different areas namely an area along the Zambezi valley in southern Tanzania, an area along the Luangwa River in northern Zimbabwe and eastern Zambia and a third area along the Shire River in Malawi (Van den Abeele, 2016). As their habitat indicates, these birds live close to rivers as they bathe several times a day. This species is popular in aviculture and display only four colour variations (Mzumara *et al.*, 2016 a; Van den Abeele, 2016). Habitat destruction and illegal trapping to sell birds to the pet trade are the biggest threats that lead to the decline in numbers and, ultimately, the CITES appendix II classification of these birds (Mzumara *et al.*, 2016 b). Poisoning of water holes in both protected and unprotected areas to control pest species responsible for crop destruction, is also a major problem this species face in Malawi (Mzumara *et al.*, 2016 b). They prefer a habitat of mopane trees (*Colophospermum mopane*) but due to deforestation, these habitats are shrinking. The birds are 14 cm tall with an orange red mask fading into a lighter colour on the bib. The mask on the back of the head gradually changes to an olive green, becoming a brighter green closer to the neck. The white ring around its eye is present as shown in Figure 2.1 h. Overall, the body is green with lighter green tail feathers and in the middle of the tail there is an orange yellow patch fading to black with a narrow yellow band surrounding it. Perrin (2009) list their natural diet as grass seeds and flowers of several tree species such as *Faidherbia* and *Erythrospermum* spp.

2.2.9 *AGAPORNIS SWINDERNIANUS*

This species is found across central and western Africa and have been observed in eight African countries. Very little is known about *A. swindernianus* as they do not survive in captivity and limited research has been conducted in their natural habitat (Van den Abeele, 2016). They are listed under the IUCN Red List of Threatened Species as "least concern" with a stable population growth due to the fact that they have an extremely large habitat range. However, Forshaw (1989) described their habitat as scarce and declining due to widespread deforestation. These birds are known to be noisy and gregarious and small flocks are typically seen very swiftly due to the fact that Swindern's (also called black collared) lovebird lives exclusively amongst tree canopies. Perrin (2009) list their diet as fruits of Strangler figs and occasionally seeds. *A. swindernianus* and *A. pullarius* coexist in the same habitat in West Africa but since their behaviours are so different no interbreeding takes place (Perrin, 2009). Forshaw (1989) describe them as the only species with a black collar and black bill at 13 cm

in height. The white eye-ring is absent in this species. Two well-described and one poorly differentiated subspecies are found namely *A. s. swindernianus*, *A. s. zenkeri* and *A. s. emini*.

2.2.10 HYBRIDS BETWEEN DIFFERENT *AGAPORNIS* SPECIES

With the exception of *A. swindernianus* and *A. pullarius* (where behaviour prevents interbreeding) the habitats of the different *Agapornis* species do not overlap and therefore the species do not interbreed in their natural habitat. This is, however, a problem in aviculture, especially in colony-style breeding systems where birds are placed in an open aviary. Mating between different *Agapornis* species can produce viable and fertile young and McCarthy (2006) list the hybrids crosses between *Agapornis* species as well as between these species and other parrots. These hybrid matings are shown in Table 2.1 and in Figure 2.2 photos of two different lovebird hybrids are shown.

Table 2.1: Viable matings between different *Agapornis* species (depicted as an x).

	A. canus	*A. fischeri*	*A. nigrigenis*	*A. lilianae*	*A. personatus*	*A. roseicollis*	*A. pullarius*	*A. taranta*
A. canus		x	x	x	x	x	x	x
A. fischeri	x		x	x	x	x		x
A. nigrigenis	x	x		x	x	x		
A. lilianae	x	x	x		x	x		
A. personatus	x	x	x	x		x	x	x
A. roseicollis	x	x	x	x	x		x	
A. pullarius	x				x	x		x
A. taranta	x	x			x			

There are also reports of viable offspring hatching from a mating between an *A. canus* male and *Melopsittacus undulatus* (budgerigar) female, *A. nigrigenis* male and budgerigar female, *A. personatus* male and budgerigar female and *A. roseicollis* male and budgerigar female (McCarthy, 2006). Hybrids are not generally accepted at auctions and shows (Van den Abeele, 2016). Even though it is beyond the scope of this study, it is important to identify Single Nucleotide Polymorphisms (SNPs) linked to specific species and hybrids in order to develop a screening test that can aid breeders to distinguish species, purebreds and hybrids.

a b

Figure 2.2: Photos of two *Agapornis* hybrid birds.
A hybrid of *A. roseicollis* and *A. nigrigenis* (a) showing that the black face of *A. nigrigenis* is absent but the white skin around the eye is present. Photo b depicts a hybrid between *A. lilianae* and *A. fischeri* birds shows the plumage on the neck is almost yellow, unlike the green colour of either species (Photo courtesy of Mr D. Van den Abeele).

Several studies have been conducted on the natural lovebird populations and the threats these birds are faced with. In conclusion it was found that reduction of their natural habitat due to agricultural activity and trapping or poaching are the main threats to all species (Tekalign & Bekele, 2011 (on *A. taranta*), Mzumara *et al.*, 2016 a & b (on *A. lilianae*), Egwumah & Iboyi, 2017 (on *A. pullarius*) and Perrin (2012) (all species)). In addition, Mzumara *et al.*, 2016 b found that poisoning of birds in Malawi is a common practice as these birds destroy crops.

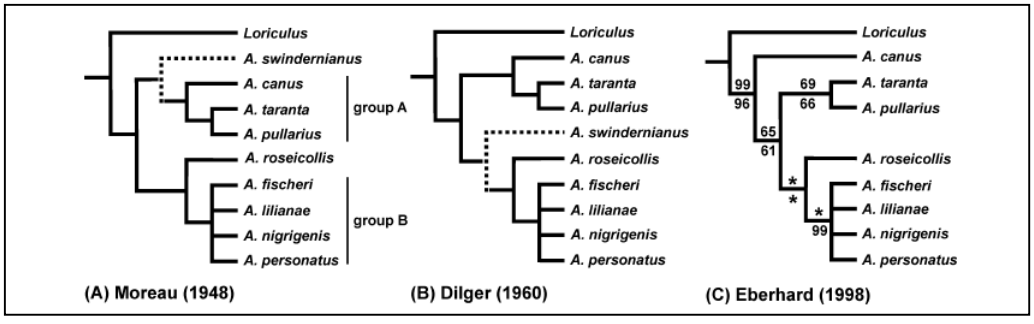
2.3 TAXONOMY AND PHYLOGENETICS

Stidham (2009) described the discovery of a fossilised lovebird humerus at the Plio-Pleistocene location of Kromdraai B in Gauteng, South Africa. The age of the fossil is estimated as early Pleistocene and therefore less than 1.95 million years old. Due to the smaller size of the fossilised humerus, compared to the same bone of extant species found today, this bird probably belonged to an extinct species. In a later study, Manegold (2013) described the discovery of yet another extinct *Agapornis* species in South Africa, 110 km NNW from Cape Town, and more than 1 000 km from the Kromdraai location. There are currently no wild parrots found in this area. The newly discovered *Agapornis* species was called *A. attenboroughi* (dedicated to Sir David Attenborough in recognition of his love of birds) and is believed to have lived in the early Pliocene time (Manegold, 2013).

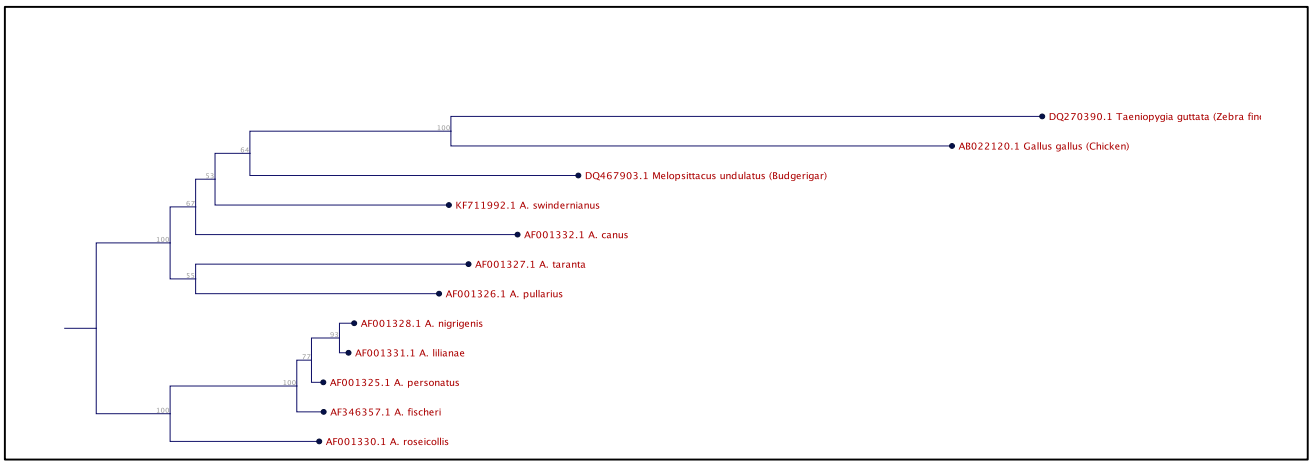
Moreau (1948) classified the nine extant lovebird species into three taxonomic groups namely Group A consisting of *A. canus, A. pullarius* and *A. taranta* and Group B consisting of *A.*

fischeri, A. personatus, A. lilianae and *A. nigrigenis*. *A. roseicollis* was considered a sister taxon of group B and *A. swindernianus* formed a sister clade with Group A. Moreau's classification was based on the young's feather colour, juvenal plumage, white skin around their eyes, being sexually dimorphic, the presence of a black bar on the birds' central tail feathers, the method used to carry nest material and the shape of their nests. Twelve years later, Dilger (1960) divided the species into an "eye-ring group" (*A. fischeri*, *personatus*, *lilianae* and *nigrigenis*) with a large patch of naked white skin around the eyes; the "sexually dimorphic group" (*A. canus*, *A. taranta* and *A. pullarius*) with males and females looking dissimilar and the "transitional group" consisting of *A. roseicollis* and *A. swindernianus* (males and females are identical but without the white eye ring). Dilger (1960), Neunzig (1926) and Hampe (1957) considered the species in the transitional group as a sub-species of one of the other species.

Eberhard (1998) sequenced a portion of the cytochrome-b (cytb) mitochondrial gene of eight (with the exception of *A. swindernianus*) *Agapornis* species. The molecular evidence from this study confirmed the classification by Dilger (1960) and categorized the species as the eye-ring group, sexually dimorph group and transitional group. In a more recent study by Manegold & Podsiadlowski (2014) a museum sample of *A. swindernianus* was also analysed at the cytb mitochondrial gene. They found *A. swindernianus* to be a sister clade to all other *Agapornis* species. In Figure 2.3(a) the phylogenetic relationships within the *Agapornis* genus showing the classifications of Moreau, Dilger and Ebenhard is given. In Figure 2.3(b) a phylogenetic tree was drawn using the same sequences as Manegold & Podsiadlowski (2014) had submitted onto The National Centre for Biotechnology Information (NCBI) using CLC Genomics workbench 11.0 (https://www.qiagenbioinformatics.com/). The Zebra finch, chicken and budgerigar were used as outgroups. The results indicate that *A. swindernianus* is a sister clade of the other species, as also concluded by Manegold & Podsiadlowski, 2014.



a



b

Figure 2.3: The classification of the nine *Agapornis* species.
In figure a, the classification as made by (A) Moreau (1948), (B) Dilger (1960) and (C) Eberhard (1998) is shown and in figure b the phylogenetic tree including the sequence of *A. swindernianus* as sequenced by Manegold & Podsiadlowski (2014).

2.4 PLUMAGE COLORATION

The main selection criterion lovebird breeders use to select breeding stock, is that of colour plumage variation. There are at least 30 variations found amongst the eight domesticated species and all are inherited. Most of the *Agapornis* species have at least one colour variation that is found infrequently in wild populations and, due to selective breeding, frequently in domesticated populations. Reports on wild lovebirds with colours other than the wildtype coloration is scarce but Warburton & Perrin (2005) observed one yellow and three light-green *A. nigrigenis* birds in the mid Machile River region between May 1999 and May 2000. Van den Abeele (2016) identified a wild-caught double factor[1] green *A. roseicollis* in Namibia. Some of the earliest reports are found in the 1932 Proceedings of the Zoological Society of London (Seth-Smith, 1931) that describes and documents the first wild-caught *A. personatus* male. Some of these variations have been known to breeders for almost a century. Toerien (1950) reports a lutino *A. roseicollis* bred from a wild caught population from Namibia, whereas Delacour (1942) reports the same variation in *A. fischeri* and Moreau (1948) in *A. lilianae*. The mode of inheritance of these variations have been determined through test matings and breeders' records and all follow a Mendelian inheritance pattern (Van Den Abeele, 2016). Despite the economic value of these traits, the genes and polymorphisms linked to these

[1] The terms "single factor" and "double factor" refer to colour variations that are inherited as incomplete dominant traits (refer to Table 2.2 for a list of such variations). "Single factor" is a bird that is a heterozygote of the phenotype and is of a slightly darker colour (green or blue) than the wildtype, but a lighter shade than a double factor. "Double factor" refers to a bird that is homozygous for the trait and these birds are darker than heterozygous (and wildtype) birds. (Van den Abeele, 2016; Hayward, 1979). Refer to section 2.4.4 and Figure 2.5 for examples.

variations in lovebirds have not yet been identified. Examples of some of these colour variations are shown in Figure 2.4.

Figure 2.4: Some of the *Agapornis* colour variations.
(All photos: Mr. Dirk Van den Abeele). The first four photos (a-d) are all of *A. roseicollis* birds - (a) Opaline single factor (heterozygote with incomplete penetrance) blue, (b) orange mask, (c) aqua and (d) opaline green. Photos e and f are of *A. personatus* birds – (e) double dark factor (homozygous for the variation with incomplete penetrance) violet single dark factor blue and pastel green (f). Photo g shows an *A. fischeri* turquoise bird whereas photo h is that of an *A. lilianae* non-sex-linked ino green bird.

There are six known mechanisms that birds utilize to express colour in their plumage, beaks, feet and combs – melanins, carotenoids, porphyrins, psittacofulvins, structural barbs and structural barbules (Stoddard & Prum, 2011). In parrots, melanins, psittacofulvin pigment and structural barbs are the mechanisms utilised. Parrots are the only group of organisms known to synthesize psittacofulvin pigments in their plumage cells (Stradi *et al*., 2001; Hudon & Brush, 1990; McGraw & Nogare, 2004, Hill & McGraw, 2006). Black, brown and red-brown plumage, beak, claw and eye colours in all bird species are caused by melanin, whereas red, yellow, pink and orange colours found in parrot plumage are caused by psittacofulvin pigments (Mundy, 2006b; Hill & McGraw, 2006). Melanin is expressed in all extant bird species (Stoddard & Prum, 2011) and the melanocortin-1 receptor (MC1R) and agouti signalling

protein (ASIP) genes stand centre in expressing these pigments (Hubbard *et al.*, 2010). In addition to yellow psittacofulvin and melanin pigments, physical interactions of light waves with β-keratin rods in the feather barbs produces purple, blue and green colours (Dyck, 1971 a & b; Prum *et al*., 1994; Prum, 2006). The two mechanism work together as follows: pigments (melanins and psittacofulvins) absorb and reflect light selectively (D'Alba *et al*., 2012) and due to the structure of the feather barbs the light is reflected irregularly (Dyck, 1971 a & b; Hill & McGraw, 2006; Berg & Bennett, 2010). The barb of a green feather, for example, contains a cortex filled with yellow psittacofulvin pigment. Light is partially absorbed by the eumelanin pigment in the feather, and blue light is reflected back through the air vacuoles. It then passes through the yellow psittacofulvin pigment and through coherent scattering of the quasi-ordered β-keratin rods, the light (and ultimately the feather) appears green (Dyck 1971 a & b). Dark green feathers reflect about half the light compared to lighter green feathers (Dyck, 1971 a & b). Mundy (2006b) stresses the fact that these two mechanisms (pigment and structural changes) are most probably under independent genetic control but that no candidate genes have been associated with structural coloration control.

2.4.1 *AGAPORNIS* PLUMAGE VARIATIONS

Buckley (1987) defines a plumage colour polymorphism as the presence of a plumage colour aberration in an interbreeding population of individuals of the same sex and age. MUTAVI, a parrot breeding research and advice group located in the Netherlands (www.MUTAVI.info) conducted research trials on the structures of *Agapornis* feathers of various colorations. In Table 2.2, a summary of the different plumage colour variations per species plus their inheritance patterns, is shown. Given that none of the polymorphisms linked to these colour variations have been identified, the colour patterns will be referred to as "plumage colour variations" and not “plumage colour polymorphisms”.

There are four main categories of colour variations in *Agapornis* and a discussion with examples thereof follows below. These include variations due to psittacofulvin reduction or modification, variations due to a change in melanin, change in feather structure and changes in eumelanin. It is important to note that many of these plumage colour variations can be combined that will result in a range of plumage colours (Hayward, 1979; Van den Abeele, 2016). Berg & Bennett (2010) as well as Mundy (2018) highlights the lack of research on the genetic control of aberrant colours in caged parrots.

Table 2.2: Plumage colour variations and their inheritance patterns found amongst *Agapornis* species

	A. roseicollis	*A. canus*	*A. taranta*	*A. pullarius*	*A. personatus*	*A. fischeri*	*A. lilianae*	*A. nigrigenis*	Inheritance pattern
Dark factor	X		X		X	X	X		AID
Blue					X	X	X	X	AR
Aqua	X					X			AR
Turquoise	X		X						AR
Orange face	X								AR
Pale headed	X								AID
Opaline	X					X			SLR
Ino (NSL[#])					X	X	X	X	AR
Ino (SL[†])	X								SLR
Pastel					X				AR
Dark eyed clear						X			AR
Pallid	X								SLR
Cinnamon	X								SLR
Pale	X					X			SLR
Marbled	X								AR
Dilute	X		X				X	X	AR
Bronze fallow	X		X			X			AR

Table 2.2: Plumage colour variations and their inheritance patterns found amongst *Agapornis* species - continued

Pale fallow	X		X			X			AR
Dun fallow					X				AR
Dominant pied	X					X			AD
Recessive pied	X					X			AR
Mottle						X			Unclear
Dominant edged						x			AID
Misty	X		X			X		X	AID
Slaty						X			AID
Violet	X				X				AID
Euwing						X			AID
Faded	X					X			AR
Crested	X					X			AID
Jade	X								AR

#: Non-Sexed linked Ino
†: Sexed linked Ino
AID: Autosomal Incomplete Dominance
AR: Autosomal Recessive
SLR: Sex-linked Recessive
AD: Autosomal Dominant

2.4.2 PLUMAGE COLORATION DUE TO PSITTACOFULVIN REDUCTION OR MODIFICATION

Some of the most dramatic colour variations amongst the psittaciform order, is caused by the reduction or modification of psittacofulvin within the feather cortex. In the 1932 Proceedings of the Zoological Society of London a wild-caught *A. personatus* male was described as "the yellow coloring was absent, the areas of the body normally yellow, being whitish and those normally green, blue" (Seth-Smith, 1931). Cooke *et al*. (2017) reported that since the 19th century captive bred budgerigars displayed blue, instead of wildtype green, plumage coloration indicating a loss of psittacofulvin pigmentation. Yellow pigment in the cortex of the feather is normally combined with blue light and results in visibly green feathers. Cooke *et al.* (2017) found that due to a T>C substitution mutation in the *MuPKS* gene in budgerigars, no yellow psittacofulvin pigment is produced and therefore blue light is reflected so that the feather appears blue. Other areas of the body that would normally be red, yellow or orange due to psittacofulvins are white due to the absence of the psittacofulvin pigment (Prum, 2006; Van den Abeele, 2016). The same phenotype is observed amongst lovebirds and it is also inherited as an autosomal recessive trait, as in budgerigars. Research done at MUTAVI (unpublished, personal communication Mr Dirk Van den Abeele) has found that no yellow psittacofulvin pigment was present in blue lovebirds compared to pigment being present in wildtype green birds. More research on the *MuPKS* gene in lovebirds is needed before conclusions can be made for this genus.

In contrast to the *Blue* phenotype where no psittacofulvin is produced, birds with *Aqua* and *Turquoise* phenotypes produce a reduced amount of psittacofulvin. There are no reports of studies where the actual pigment is measured, but visibly about 50% of the normal amount of psittacofulvin is produced in *Aqua* birds and only about 40% of the normal levels of psittacofulvin is visible in birds with the *Turquoise* variation (Van den Abeele, 2016).

Orange face and *Pale headed* are two popular phenotypes in lovebird aviculture and are only found amongst *A. roseicollis* (Van den Abeele, 2016). The mechanism behind these phenotypes are poorly understood (Personal communication, Mr Dirk Van den Abeele) but for *Orange face* individuals, red plumage is changed to orange, whereas in *pale headed* individuals the psittacofulvin change from red to pink. Green plumage remains unchanged for both phenotypes. Research by McGraw & Nogare (2004) indicated that the red, orange and pink coloration are all due to psittacofulvin pigment in different concentrations and therefore these colour variations could be caused by a change in psittacofulvin pigment concentration.

2.4.3 COLOUR VARIATIONS DUE TO A CHANGE IN MELANIN

Mundy (2006a) and McGraw (2006) noted that most colour variations are caused due to changes in melanin distribution. However, little to no molecular research has been conducted on parrot species regarding the genetic basis of these variations. Most of the common types of bird plumage colour variations can be greatly attributed to a lack of melanin in the feathers. Examples of plumage coloration variations due to a change in melanin in *Agapornis* is briefly given in Table 2.3.

Table 2.3: Colour variations due to a change in melanin in *Agapornis* species

Variation	Phenotype	Reference
Opaline	Rearrangement of colour pigments causing unusual colour patterns.	Hume & van Grouw, 2014
Ino sex linked (*ino-SL*)	Green > Yellow Blue > White Dark red eye Pale pink feet	Gunnarsson *et al*. (2007), Van den Abeele, 2016.
Ino non-sex linked (*ino-NSL*)	Green > Yellow Blue > White Bright red eye	Van den Abeele, 2016.
Pastel	Paler green colour.	van Grouw, 2013; Van den Abeele, 2016
Dec	Primarily yellow with a green hue.	Van den Abeele, 2016
Pallid	Hatch with red eyes turn darker.	Van den Abeele, 2016
Pale	Paler than wildtype.	Van den Abeele, 2016
Dilute	Very light green / yellow colour.	Van den Abeele, 2016
Misty	Heterozygotes: dull colour, homozygotes: resemble double factor green variant.	Van den Abeele, 2016
Bronze-, *Pale-* and *dun-fallow*	Some eumelanin still present, has red eyes.	Van den Abeele, 2016
Cinnamon	Black feathers are brown, paler green, lighter blue rump.	Hayward (1979), Tsudzuki (2008)
Dominant and *recessive Pied* (*piebald*)	Green > Yellow Blue > White Not expressed in all feathers.	Van Grouw (2013), Van den Abeele, 2016; Guay *et al*. 2012; Hayward, 1979.
Marble	Feather tips have a darker edge.	Yakovlev *et al*. (1975); Van den Abeele, 2016.
Dominant edged	Heterozygotes: clear darker edges on wingtips, homozygotes: yellow with light grey flight feathers and a green hue on wings.	Van den Abeele, 2016

One example of a change in melanin is pale fallow, where some eumelanin is still present in the feathers and the eyes of the bird are red. Figure 2.5 shows an *A. taranta* bird with this variation.

Figure 2.5: A pale fallow *A. taranta* bird.
Almost all green over its body is absent and plumage is a yellow colour. The red on its head is still present. Photo: Mr Dirk Van den Abeele.

2.4.4 COLOUR VARIATIONS DUE TO A CHANGE IN THE FEATHER STRUCTURE

The "*dark factor*" phenotype causes the spongy zone of the feather to decrease in size. Due to this reduction only about half the light is reflected throughout the visible spectrum compared to light green feathers (Dyck, 1971 a & b). The inheritance of this phenotype is incomplete dominance and three phenotypes are observed. Heterozygote birds (also called "*single dark factor*") are slightly darker (a khaki green) than the wildtype and birds that are homozygous for the mutation ("*double dark factor*") are found to be an olive green colour (Hayward, 1979; Van den Abeele, 2016; Unpublished data, MUTAVI). This phenotype is observed in five of the *Agapornis* species as seen in Table 2.2. It is also found amongst the budgerigar resulting in the same colours (Harris, 1979). Should no psittacofulvin pigment be produced the same effect is seen in blue birds. Therefore, two variations are combined in these birds (dark factor and blue). This is shown in Figure 2.6 where the difference between two *A. personatus* blue birds are shown.

Figure 2.6: Single and double factor blue *A. personatus* birds.
The bird on the left is a single dark factor blue bird and the one on the right is a double dark factor blue bird. Note the lighter shade of blue of the bird on the left compared to the one on the right. (Photo curtesy of Mr Dirk Van den Abeele.)

Research projects conducted at MUTAVI, discovered that lovebirds with the *slaty* phenotype had a change in the keratin in the feathers from a brownish horn colour to a glassy see-through colour (Personal communication: Mr. Dirk Van den Abeele) and plumage colour change to a steel blue colour. A similar phenotype, called *slate*, is commonly found amongst budgerigars (Elliott, 2013) and is caused by changes in the feather structure and inherited as a sex-linked recessive trait. Through test matings of *A. fischeri*, *A. nigrigenis* and budgerigars, it was determined that this phenotype is inherited as an incomplete dominant trait. However, there is visually no difference between a heterozygote and a homozygote for the mutation and therefore genotypes cannot be determined based on phenotype alone (Van den Abeele, 2016). Further research is necessary to identify the genes and polymorphisms linked to these phenotypes and to confirm whether the *slaty* and *slate* phenotypes are caused by the same polymorphism.

The *violet* phenotype develops due to a structural change in the spongy zone of the feather. It is however only visible when it is combined with the blue phenotype. If psittacofulvin is absent in the cortex (thus a blue bird), combined with the structural change in the spongy zone, the bird appears violet. The same phenotype is found in budgerigars with the same underlying inheritance pattern (Harris, 1979; Van den Abeele, 2016). There is a difference in appearance between a heterozygote and homozygote for the variation in both of the species.

2.4.5 COLORATION DUE TO CHANGES IN EUMELANIN

In 2017 the "*Jade*" phenotype was officially accepted as a plumage colour aberration in *A. roseicollis*, and is inherited as an autosomal recessive trait. This phenotype has not been described in the literature and all observations are from personal communication with Mr. Dirk Van den Abeele. Interestingly, breeders of these birds observed that females are lighter than

males. Birds with this variation appears to have a visible reduction of about half of the eumelanin in their wing feathers and more so over their body, resulting in a yellow colour with a darker hue, as can be seen in Figure 2.7. There is a greater eumelanin reduction in the core of the wing feathers which results in a slight edge on the feathers. The bird's rump is much lighter and the legs are of normal coloration, but the nails are darker. The mask of these birds are of normal coloration.

Figure 2.7: The Jade phenotype as observed in *A. roseicollis*.
Photo curtesy of Mr Dirk Van den Abeele.

The *Euwing* pattern is a colour variation only found amongst A. *fischeri* causing darker wings and an overall different body colour. The cause of the phenotype is still unclear but is inherited as a dominant trait with incomplete penetrance (Van den Abeele, 2016). Veeckmans (2016) notes that the *euwing* mutation "dulls" the black eumelanin pigment in the body feathers and on the mantle in heterozygote birds. In homozygous affected birds, the effect is much more prominent with additional eumelanin found in the wing covert feathers, making the bird appear darker. By combining a bird with both the *euwing* variation as well as the *opaline* variation, breeders can achieve spectacular colorations. The euwing, opaline and a combination of the two variations in *A. fischeri* are shown in Figure 2.8.

a b c d

Figure 2.8: Euwing and opaline variations in *A. fischeri*.
(All photos: Mr. Dirk Van den Abeele). Picture (a) shows a wildtype *A. fischeri* bird, in picture (b) a single factor euwing green bird is shown, in picture (c) a photo of an opaline green bird is given and in picture (d) the combination of single factor euwing green and opaline is shown.

2.5 *AGAPORNIS* BREEDING

Worldwide, there is a keen interest in lovebird breeding. Membership of the world's largest breeding association - The Belgium Agapornis Breeders Association (BVA-International) - is approximately 5000 members (Personal communication, Mr Dirk Van den Abeele) and other societies across Europe, Africa, Australia, the United States of America, South America and Asia (Van den Abeele, 2016) exist with various member numbers.

Birds are not registered at a registration body in South Africa, and thus limited information is available regarding the number of lovebirds that are bred each year in this country. Approximately 25 000 lovebird leg rings are sold by the Parrot Breeders Association of South Africa (PVSA) but there are other ring suppliers as well (Personal communication: PVSA). There is also no registry of how many South African lovebird breeders or breeding birds exist. Each breeder can have any number of birds and since lovebird eggs hatch 21-24 days after laying and birds can hatch several clutches every year, many birds are hatched annually.

In lovebird breeding worldwide the birds of a higher "quality", which is dictated to a large degree by the pedigree of the bird and the colour variations the bird display or might be a heterozygote for, are sold for a premium compared to a wild type bird (Personal communication, Mr Dirk Van den Abeele). Despite the colouration and pedigree influencing the sale price of a bird, no confirmation of pedigree or other data is required to sell, export or import a bird. Test matings need to be performed by a buyer in order to determine if a bird is indeed a heterozygote for a particular trait and is mostly only confirmed a few months or even years after a sale is made (Personal communication, Mr Dirk Van den Abeele).

CITES keeps records of all imported and exported live animals under Appendix I, II and III (therefore all *Agapornis* species except *A. roseicollis*) (https://trade.cites.org/en/cites_trade/). Import or export permits are not required for species not under CITES Appendixes and therefore very little data is available for *A. roseicollis* imports and exports. Figure 2.9 a-g indicates all *Agapornis* birds (per species) for *A. fischeri, A. canus, A. lilianae, A. nigrigenis, A. personatus, A. pullarius* and *A. taranta* exported from South Africa between 2012 and 2016 (per year). The total number of *Agapornis* live birds (excluding *A. roseicollis*) exported globally (thus exported to and from any country) between 2012 and 2016 is shown in Figure 2.9 h. It is unclear if no birds were exported for *A. fischeri*, *A. canus*, *A. nigrigenis*, *A. pullarius* and *A. taranta* between 2015-2016, or if no data was made available by CITES. Most of these birds were exported for the pet market. The numbers correlate with the popularity of the respective species in the pet market. A 2012 poll by Ornitho Genetics completed by 530 voters showed that *A. roseicollis* (27%) was the species most widely bred with, followed by *A. fischeri* (24%) and in third place *A. personatus* (22%). In 2017 the same poll was completed by 643 voters placing *A. fischeri* in the first place (41%), *A. personatus* second (24%) and *A. roseicollis* in third place (15%) (https://www.ornitho-genetics.info/?s=poll). *A. fischeri* and *A. personatus* were the two species exported the most according to the 2012 – 2016 CITES data. The popularity of a species in aviculture could drastically influence their numbers in the natural habitat as traders will trap these birds in order to sell them. Perrin (2012) reports that *A. roseicollis*, *A. personatus* and *A. fischeri* are all heavily traded whereas *A. lilianae* are traded moderately and *A. nigrigenis* to a lesser extent. Trade and trappings are banned in several African countries which should have a positive effect on the natural populations' growth. As no data is available for *A. roseicollis* that comparison can unfortunately not be made. Figure 2.9 also shows a decline in the number of birds that were exported, which could be accounted for by the outbreak of bird flu and stricter import legislation to the European Union (Barnes, 2017).

Figure 2.9: *Agapornis* birds exported between 2012 and 2016.
a-g: Number of *Agapornis* individuals per species exported between 2012 and 2016.
h: Number of *Agapornis* individuals exported globally between 2012 and 2016.

Although the lovebird breeding industry is not the largest amongst parrot species, it is actively growing in many parts of the world. The development of genetic screening tests is therefore of paramount importance. As with other companion animals like dogs and cats, lovebirds aren't selected based on production traits (such as carcass weight or growth) but rather based on appearance. Plumage colouration is considered the trait most widely used to select birds as breeding stock. As discussed in the previous section, most of these variations are inherited as Mendelian traits. Currently none of the causative mutations linked to colour variations in lovebirds have been confirmed or identified. Therefore, breeders use the pedigree of a bird to predict the likelihood of it being a heterozygote of a specific variation. This implies that the pedigree records of a bird need to be highly accurate and complete before a sale is made.

Pedigree records are currently not routinely tested for any lovebird species in any country. A handful of service laboratories globally (e.g. https://www.genomia.cz/en/birds and https://www.animalgenetics.us/Avian/DNA_Typing/Avian-DNA-Typing.asp) offer DNA profiling and parentage verification for birds and parrots in particular, but these are all microsatellite based tests and not species specific. Since microsatellite markers often do not cross-amplify across species this could lead to poor test results. In 2014 The International Society of Animal Genetics (ISAG) suggested a panel of 16 microsatellite markers for pigeon genotyping but no panel of SNP markers or a parrot panel has been suggested as yet (https://www.isag.us/docs/B_Pigeons_OthersSpecies2014.pdf). This is in direct contrast to the direction that other breeding animals' pedigree recording is moving, where SNPs are routinely being used to verify parentage (Heaton *et al.*, 2014; Qiu *et al.*, 2016).

In order to develop such a SNP-based parentage verification panel, SNPs have to be identified throughout the genome. By sequencing, assembling and annotation of the *de novo* genome, SNPs, genes and polymorphisms can be identified with greater ease than without a reference genome.

2.6 AVIAN GENOMICS AND *DE NOVO* GENOME SEQUENCING

Prior to Next Generation (NGS) Sequencing, genomes were analysed using linkage maps. The chicken (*Gallus gallus domesticus*) is considered one of the avian model organisms and Groenen *et al.* (2000) compiled a consensus linkage map using Sanger sequencing reads for this organism. It was concluded that the chicken genome map spans 3800 cM and contains 2n=78 chromosomes consisting of micro- and macrochromosomes (Groenen *et al*., 2000), and 1889 loci. The first draft of the chicken genome was sequenced using NGS in 2004 (Hillier *et al.*, 2004) and in 2017 the fifth version (Gallus_gallus-5.0) of the genome was released (Warren *et al*., 2017). From the fifth version they concluded that the genome was 1.21 Gb in

size and annotated 26 640 genes (19 119 protein-coding, 6839 non-coding and 627 pseudogenes). The chicken belongs to the order of Galliformes (non-passerine order) and the linkage map of the first passerine species, the zebra finch (*Taeniopygia guttata*), was completed in 2008 (Stapley *et al.*, 2008). The first NGS draft genome of this species was completed in 2010 (Warren *et al.,* 2010).

The second decade of the twenty-first century saw a significant drop in cost of genetic sequencing and led to a higher number of *de novo* non-model organism genomes being sequenced. The Genome 10K Project was conceptualized in 2009 (Genome 10K Community of Scientist, 2009) and aimed to sequence 16 203 vertebra species without reference genomes spanning mammals, birds, non-avian reptiles, amphibians and fish. From this project the Bird 10K project was initiated. On 3 June 2015 it was announced that the team aimed to sequence more than 10 500 extant bird species over the next five years (Zhang *et al.,* 2015). Four phases were proposed – Phase 1, to sequence all 34 bird orders; Phase 2, to sequence 300 bird families. Both of these phases have been completed. As per their website on 5 July 2017 (last update) (https://b10k.genomics.cn/progress.html) Phase 3 and Phase 4 are still ongoing and entails specimen and trait data collection of around 2 500 species (Phase 3) and sequencing the remaining 10 500 species (Phase 4).

De novo genome assembly is defined as merging and ordering short DNA fragments (reads) that were sampled from a genome, into the correct order to reconstruct the genome, for organisms without an assembled reference genome (Paszkiewicz & Studholme, 2010; Earl et *al.*, 2011; Baker, 2012; Wajid & Serpedin, 2014). The first step in this process is to sequence the complete genome of the organism to produce reads that could be assembled into the complete genome. There are various sequencing platforms available including Illumina (e.g. HiSeq); PacBio and Roche 454. Thereafter, the reads are assembled in two main steps namely constructing contigs and constructing scaffolds. The assembly of these reads is where the real challenge of the process lays as read lengths and read counts differ and error profiles are found (Bradnam *et al.*, 2013). The development of a pipeline to order the raw reads into a complete and accurate genome sequence requires experience and a suitable assembler. Some of the assemblers available include SOAPdenovo2 (Lou *et al.*, 2012), ABySS (Simpson *et al.*, 2009), ALLPATHS (Butler *et al.*, 2008), SGA (Simpson & Durbin, 2012) and Velvet (Zerbino & Birney, 2008). After genome assembly, gene elements are named in the process called annotation. Once a reference genome is assembled and annotated, more information on the size of the genome, number of genes, gene sequences, SNPs location etc. can be gathered which allows for the development of, amongst others, genetic screening tests.

A couple of *de novo* avian genomes were sequenced in the recent past. The methods and results of these projects are discussed in more detail in Chapter 3 (see Table 3.1). These include four parrot species (scarlet macaw (*Ara macao*) (Seabury *et al.*, 2013), the Puerto Rican parrot (*Amazona vittata*) (Oleksyk *et al.*, 2012), the blue fronted Amazon (*Amazona aestival*) (Wirthlin *et al*., 2018) and three different assemblies of the budgerigar genome (Ganapathy *et al.*, 2014), as well as 48 bird genomes (which also included the budgerigar) sequenced as part of the Bird10K project (Zhang *et al.,* 2014 a and b). These studies, especially those of the parrot species, gave some insight on what to expect regarding the lovebird genome size and the number of genes as well as metrics to assess the quality of the assembly such as the N50 lengths. It also shows a wide range of sequencing platforms (e.g. Illumina, Roche and PacBio) and assemblers (e.g. SOAPdenovo, AllPaths and ABySS) that were applied.

The first "genomic" study on *Agapornis*, focussing on karyotyping of the genome of *A. roseicollis* was conducted by Nanda *et al*. (2007). This study revealed that this species has a diploid number of 48 macro-chromosomes. In the same study the budgerigar (*Melopsittacus undulates*) was found to have 2n = 62 macro-chromosomes. The lower number of *Agapornis* chromosomes might indicate a fusion of ancestral micro-chromosomes (Nanda e*t al.*, 2007). Most of the genomes mentioned here and discussed in Chapter 3 were not assembled up to chromosome level. Two of the genomes included in the Zhang study (pigeon, *Columbia livia* and Peregrine falcon, *Falco peregrinus*) were assembled to chromosome level in a later study by Damas *et al.* (2017). The remaining genomes were only assembled to scaffold level. The chicken (Warren *et al.*, 2017), zebra finch (Warren *et al*., 2010) and turkey (Dalloul *et al*., 2010) are a few of the species of economic importance that has been assembled up to chromosome level. None of the parrot genomes were sequenced up to chromosome level.

No study to sequence, assemble and annotate the genome of any of the lovebird species has been conducted thus far and no reference genome exists for any of the nine *Agapornis* species. The lack of a reference genome for lovebirds as well as the need for genetic screening tests such as a SNP-based parentage verification test highlights the need to sequence the reference genome of one of the lovebird species. Sequencing up to scaffold level should be adequate to identify SNPs for this purpose. SNP identification for parentage verification should not be influenced by the gene annotation since the preferred location of parentage-SNPs is non-coding DNA.

2.7 SNP DISCOVERY

SNPs are identified or discovered when a read of an individual of the same species is compared with the reference genome and a specific nucleotide differs between the two (Kumar *et al.,* 2012). Seeb *et al*. (2011) categorizes two techniques to discover SNPs - chain-termination sequencing (Sanger sequencing) and NGS and each has several methods to identify SNPs. One of the NGS-based methods, whole genome resequencing (WGR) (see below for examples), had shown to be highly successful in previous studies in identifying SNPs.

Whole genome resequencing (WGR) entails sequencing the whole genome of an individual at a low coverage (<50x coverage) with the aim to discover variants by comparing the two reads to the same species' reference genome (Stothard *et al.*, 2011; Kardos *et al.*, 2016). One of the programs available to use as a variant caller during WGR is Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). GATK is executed in the command-line interface and gives clear set best-practices guidelines. In brief, the reads of the individual sequenced at a lower coverage are mapped to the reference genome and following the pipeline set out by GATK, variants are called. This pipeline is discussed in full detail in Chapter 4.

In one such WGR study, Flemming *et al.* (2015) compared the re-sequenced data of two highly inbred (70 generations of sib-mating) chicken lines – Fayoumi and Leghorn – with that of the Red Jungle Fowl reference genome. Genome coverage was not given. Sixteen birds per line was re-sequenced and aligned to the reference genome and SNPs throughout the genome was identified using GATK. Over 4 million SNPs were identified for both lines and over 72% novel SNPs were identified for both lines. Even in highly inbred avian samples, this method was accurate in identifying SNPs. In a bovine study, one Heugu bull (Korean Black cattle breed) (Choi *et al.*, 2013) was re-sequenced (coverage not given) and mapped against the bovine reference genome (UMD 3.1) and over 6 million SNPs were identified using GATK of which 1.8 million SNPs were novel. The third study involved the identification of SNPs for inclusion in a parentage verification panel (Kaiser *et al.*, 2017 on the black throated blue warbler, *Setophaga caerulescens*). Instead of a reference genome, they assembled the transcriptome and mapped the reads of 11 individuals to this assembly and discovered 122 680 SNPs using GATK. In a study by Schaefer *et al.* (2017) 153 horses representing 24 breeds were sequenced at a medial coverage of 13x. The reads were mapped to the equine reference genome (EquCab 2.0) and two different variant callers (GATK and SamTools (Li *et al.*, 2009)) were used. GATK called ~31 million SNPs and SamTools ~26 million and of these ~22 million overlapped.

Van der Jagt *et al*. (2018) examined the use of GATK for discovering variants in the 1000 Bull Genome project for cattle. This project aims to build a database of sequence variants for bulls. They compared the use of SamTools v1.3 (Li *et al*., 2009) with that of GATK v3.5 on 134 animals sequenced at >20x coverage. They concluded that GATK performed equally well compared to SamTools in discovering SNPs, and outperformed SamTools in insertion and deletion (indel) discovery.

From these studies it can be concluded that GATK is robust enough to accurately identify millions of SNPs throughout the genome. Seeing that a minimum of 100 SNPs is needed for a parentage verification panel (Strucken *et al.*, 2016 and see next section), the use of GATK and the resequencing method should yield an adequate number of SNPs for this application. Since one of the prerequisites of SNPs included in a parentage verification panel is that they are not found in an indel (Heaton *et al*., 2014), GATK will be a good choice as variant caller to apply in the lovebird parentage study.

2.8 COMPILING A PARENTAGE VERIFICATION PANEL

Colour variations is the trait most widely used to select lovebirds as breeding stock. Most of these 30 colour variations are inherited as Mendelian traits, but the genes and mutations linked to the variations are yet to be discovered. Therefore, should a chick hatch with wildtype coloration, breeders use pedigree data (including the parents' coloration) to predict the probability of that bird being a heterozygote of a specific variation. Pedigree information must therefore be accurate and complete, but this is not always the case. A SNP-based parentage verification test that a breeder or buyer of a bird can routinely request before a sale is made, is the solution to incomplete and inaccurate pedigree information.

The most popular molecular markers applied in animal breeding and conservation include Random Amplification of Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Restriction Fragment Length Polymorphisms (RFLP), microsatellite markers and SNPs (Beuzen *et al.*, 2000; Vignal *et al.*, 2002). In the last decade SNPs have surpassed all other markers in popularity and is now widely applied in, amongst other applications, parentage verification (Rohrer *et al*., 2007; Strucken *et al*., 2014). SNPs are distributed at an abundant rate of around 1 SNP in every 300bp of genomic DNA in coding and non-coding areas of the genome (Olsen *et al.*, 2011; Seeb *et al*., 2011). These markers are relatively easy and inexpensive to genotype and data exchange between laboratories is easy and more accurate than microsatellite markers (Garvin *et al*., 2010; Olsen *et al.*, 2011; Strucken *et al.*, 2016). Data exchange is especially important for parentage analyses where the parent animal might have been previously genotyped (often at another laboratory) and

the profile of this animal can be used to confirm the relationship. On the down side, SNPs are bi-allelic - therefore more SNPs than multi-allelic microsatellites are required to have the same exclusion power in a parentage verification panel (Olsen *et al.*, 2011).

Despite the wide use of SNPs as a selection tool in production animals and poultry (Matukumalli *et al.,* 2009; Heaton *et al*., 2014; Gheyas *et al.*, 2015) as well as pet breeding (Fels & Distl, 2014; Qui *et al*., 2016), the application of SNPs in aviculture, and especially parrot breeding, has been very limited. In the studies where SNPs have been identified for application of parentage or tractability in bird species, it has mainly been for conservation purposes – see Weinman *et al*., (2015) on superb starlings (*Lamprotornis superbus*) and Kaiser *et al.*, (2017) on the black throated blue warbler (*Setophaga caerulescens).* The only parentage verification panels currently available for parrots are microsatellite based. Coetzer *et al.,* (2017) developed a panel of 16 microsatellite markers to verify parentage of the Cape parrot (*Poicephalus robustus*) and Jan & Fumagalli, (2016) identified 106 microsatellites that was amplified amongst seven different parrot species and could be used for individual identification. As mentioned earlier, the ISAG recommended avian parentage verification panel is microsatellite based and was developed specifically for pigeons (https://www.isag.us/docs/B_Pigeons_OthersSpecies2014.pdf). In the niche field of lovebird breeding and genetics, no studies focussing on identifying SNPs, or any genetic marker, for individual or pedigree identification for any of the species in this genus, currently exist. The lack of a reference genome has made the identification of genetic markers problematic.

From previous studies where parentage verification panels were compiled, a list of the recommendations to consider when selecting SNPs for a parentage verification panel are summarised in Table 2.4. The first five criteria can be addressed by selecting the correct SNPs using a variant caller such as GATK. The last four recommendations can be addressed once the panel of SNPs have been were assessed in a large population.

Table 2.4: Recommendations for selecting SNPs to include in a parentage verification panel.

Criteria	Reference	Implication for this study
Location: physical location of each SNP and linkage disequilibrium (LD).	Heaton *et al*. (2014); Liu *et al., 2016;* Talenti *et al.,* 2016.	
Must be mapped to reference genome.	Liu *et al*., 2016; Heaton *et al*., 2014.	
Only two alleles detected.	Heaton *et al*., 2014	Ensures Mendelian inheritance.
Not part of an indel.	Heaton *et al*., 2014	All indels must be excluded during variant calling.
Consistent Mendelian inheritance pattern	Heaton *et al*., 2014; Talenti *et al*., 2016	Use relatives to ensure SNPs discovered are inherited.
Minor allele frequency (MAF)	Strucken *et al.*, 2016; Heaton *et al.*, 2014*;* Talenti *et al.,* 2016.	Values should be >0.3.
Mean heterozygosity of the panel and Observed Heterozygosity (H_O) of each SNP	Kaiser *et al*., 2016; Morin *et al.*, 2004.	Values should be >0.3.
SNPs should be in Hardy-Weinburg equilibrium (HWE).	Liu *et al., 2016;* Talenti *et al.,* 2016.	
Low genotypic error rate	Talenti *et al.*, 2016; Strucken *et al.,* 2016.	Increases accuracy of panel and ensures inheritance pattern is correct.

Important factors that will influence the accuracy of a panel and the number of SNPs needed in a panel to accurately verify parentage includes the Exclusion Probability (P_E) of the panel, heterozygosity of the individual SNPs (H_O) and panel (mean H_O), MAF and genotyping error rates.

Different mathematical models to verify identity or relationships between two individuals include categorical and fractional allocations, genotype reconstruction and, the simplest

theory, exclusion based on genotypic mismatches (Heaton *et al.*, 2014). Exclusion relies on the exclusion probability (P_E) of a set of genetic markers by using the formula described by Jamieson & Taylor (1997). This formula takes into account the genotypes of the two individuals compared, the number of loci compared and the allele frequency in a population (Jamieson & Taylor, 1997). The exclusion principle (also called "opposing homozygotes" or oph) relies thereon that a conflict of alleles (or genotypic mismatch) at a specific locus between two individuals excludes the possibility of the supposed relationship between the two individuals (Baruch & Weller, 2008). Parentage allocation programs such as Cervus 3.0 (Marshall *et al.*, 1998) uses this theory to verify relationships.

Two studies on two different bird species (Weinman *et al.* (2015) on superb starlings (*Lamprotornis superbus*) and Kaiser *et al.* (2017) on the black-throated blue warbler (*Setophaga caerulescens*) were conducted to compare the use of SNPs and microsatellites as parentage verification markers. In the warbler study, 97 SNPs were included in the final panel and in the starling study 103 loci were included. In both avian studies the effect of a high mean heterozygosity (H)[2] value of the panel dictated which and how many SNPs were included. In the warbler study, a panel of 40 SNPs with a mean observed heterozygosity (H_O) of 0.37 had the same non-exclusion power ($NE_P = 1\text{-}P_E$) ($NP_E = 1.9x10^{-2}$) as the full panel of 97 SNPs. The starling results indicated that less than 60 SNPs can be used to accurately infer parentage with a non-exclusion probability of $8.4x10^{-5}$, given the panel has a mean observed heterozygosity value between 0.4 and 0.5. All SNPs had a H_O value greater than 0.3 and a genotyping accuracy rate of ≥0.99%. Weinman *et al.* (2015) concluded that the number of SNPs required is linked to the average heterozygosity of the set of loci tested and more heterozygous SNPs would require less loci. Similarly, Morin *et al.* (2004) found that 40-100 SNPs with a H_O of 0.3 were as powerful in excluding incorrect parents as 7-14 microsatellite markers with a H_O of 0.75. Therefore, the number of SNPs to be included in the final panel is to a large extent dictated by the combined heterozygosity of the panel as well as each SNP's H_O.

Heaton *et al.* (2014) concludes that several factors may reduce successful parentage allocations, including poor assay designs, ineffective genotyping platforms and poor DNA quality. Many of these factors can be addressed by using good quality samples, a DNA

[2] Heterozygosity is a measure of genetic variation in a population. Expected Heterozygosity (H_E) can be calculated as $HE = 1 - \sum_{i=1}^{n} pi^2$ H_E is however not very sensitive to additional variation because the upper limit (1) is the same for any number of alleles, especially in highly variable markers such as microsatellites. In most outbred populations the Observed Heterozygosity (H_O) is similar to the H_E, and is useful in populations where genotypic frequencies may not be close to Hardy-Weinberg equilibrium. It is calculated as $HO = \sum_{i<j}^{n} Pij$ where P_{ij} is the estimated frequency of genotype ij. (Hedrick, 2000). Mean heterozygosity is the combined heterozygosity of all SNPs across the panel.

extraction kit designed to extract DNA from the specific tissue, proper SNP selection and discovery as well as assay design and using a good genotyping platform. Heaton *et al.*, 2014 also stresses that the P_E value of a panel is directly influenced by the Minor allele frequencies (MAF) of the SNPs as well as the genotyping error rate since a high error rate will lead to more mismatches and a low MAF value indicates that many of the genotypes will be identical (and homozygous) so exclusion isn't possible (Heaton *et al.*, 2014; Talenti *et al.*, 2016). Liu *et al*. (2016) concludes that by using a larger number of SNPs, the allocations become less sensitive to MAF values and genotyping errors.

Strucken *et al.* (2016) reviewed different metrics to determine the number of SNPs to use in a study. They started by reporting that the International Society of Animal Genetics (ISAG) recommended at least 100 SNPs to be included in a panel for accurate parentage verification, given a genotyping error rate of 1%. They genotyped 2 087 cattle from three different breeds at 33 159 SNPs, and 1 471 sheep, from two breeds, at 48 599 SNPs. Strucken *et al.* (2016) stated that the exclusion probability formula doesn't take MAF and the genotypic error rate into account. Therefore, even though a non-parent can be correctly excluded as a parent, a true parent could also be excluded. This could be due to genotyping errors that causes genetic mismatches between parents and offspring. Mismatches are the most important factor taken into account by the exclusion probability formula. The number of genotypic mismatches of a true parent-offspring pair should, in theory, be zero due to Mendelian inheritance, but since genotyping errors are inevitable, this value, in reality, will rarely be zero. If the error rate is high, leading to false genotypic mismatches, P_E will be incorrectly lower. They concluded that the number of markers, rather than the quality of markers is a determining factor of success of a parentage test. If more markers are included the effect of the error rate is decreased and the probability of correct assignments is increased. They recommend using at least 200 SNPs when one parent's genotype is known and should the MAF scores of the SNPs be lower than 0.15, the number of markers should be increased even further.

The number of SNPs in the final panel will depend on the heterozygosity, MAF values as well as the genotyping error rate of the individual SNPs as well as the combined panel. However, ISAG's guidelines of at least 100 SNPs and Strucken's recommendation of at least 200 SNPs in the panel should be taken into account. Less SNPs will be more economical to genotype commercially. Studies such as Liu *et al.* (2016) (on rainbow trout) and Labuschagne *et al.* (2015) (on African Penguin, *Spheniscus demersus*) proved that as little as 48 SNPs (trout) and 31 SNPs (penguins) can be used to accurately infer parentage, but false positive or negative allocations (as discussed by Strucken *et al*., 2016) should be taken into account. Therefore, a panel consisting of 100-200 SNPs with a mean H_O of >0.3 and MAF values of

>0.3, tested in a wide population of unrelated birds, and selected based on all criteria listed in Table 2.4, would be ideal.

CHAPTER THREE: SEQUENCING, ASSEMBLY AND ANNOTATION OF THE *AGAPORNIS ROSEICOLLIS* GENOME

3.1 INTRODUCTION

An assembled and annotated reference genome is not available for any of the *Agapornis* species. The aim of this study was to develop SNP-based genetic screening tests, including a parentage verification panel, for lovebirds. A reference genome from where SNPs can be identified makes the development of such tests easier. As described in Section 2.3 the eight domestic *Agapornis* species can be divided into three phylogenetic groups based on characteristics such as, amongst others, being sexually dimorphic and having a white ring around their eyes (Eberhard, 1998). *A. roseicollis* is the only species in the transitional group and displays characteristics of both the sexually dimorphic group and the white eye ring group. Therefore, the genome of *A. roseicollis* was sequenced, assembled and annotated. In this chapter we describe the genome sequencing, assembly and annotation process. Motivations for the use of certain methods (based on methods used in previous studies) are discussed here. The findings were published in a peer-reviewed article given at the end of this chapter.

3.2 MOTIVATION FOR METHODS USED IN THIS STUDY

A number of avian *de novo* genome projects have been conducted the last couple of years. These include the first phase sequencing of the Bird10K project by Zhang *et al.* (2014 a and b) where the *de novo* genomes from 48 different avian species were sequenced, assembled and annotated. These species represented all the major orders, including the budgerigar from the psittaciformes (parrots) order. Three additional parrot genomes that were previously sequenced but not as part of the Bird10K project, include the scarlet macaw (*Ara macao*) (Seabury *et al.*, 2013), the Puerto Rican parrot (*Amazona vittata*) (Oleksyk *et al.*, 2012) and the blue fronted Amazon (*Amazona aestival*) (Wirthlin *et al*., 2018). In Table 3.1, a summary of the methods used and results found during these studies, is given. This includes the sequencing platform used, coverage, size of the genome found, assembler used and metrics such as the contig- and scaffold-N50 lengths.

Table 3.1: Sequencing platforms, genome coverage, genome size, assembler and N50 lengths of *de novo* parrot and other avian genomes.

Species	Sequencing platform	Coverage	Genome size (Gb)	Assembler	Contig N50 (Kbp)	Scaffold N50 (Mbp)
Blue fronted Amazon	Illumina Roche-454	117.59x 3.1x	1.126	AllPaths-LG (Butler *et al.*, 2008) ABySS (Simpson *et al.*, 2009) Meraculous (Chapman *et al.*, 2011).	27.8	1.09
Budgerigar_v6.3	Illumina HiSeq 2000 and Roche-454	14x	1.2	Celera (Denisov *et al.*, 2008) CABOG (Miller *et al.*, 2008)	55.6	10.6
Budgerigar	Illumina and Roche-454	144.44x	1.2	SOAPdenovo2 (Lou *et al.*, 2012).	51	13.4
Budgerigar	PacBio corrected with Illumina & 454	17x	1.2	PBcR assembler (Koren *et al.*, 2012)	102	1.7
Scarlet Macaw	Roche 454 Illumina GAIIx Illumina HiSeq 2000	26x	1.11-1.16	CLC Genomics*	6.37	0.01597
Puerto Rican parrot	Illumina HiSeq 2000	26.89x	1.58	Ray (Boisvert *et al.*, 2010).	0.00684	0.0000190
48 Zhang *et al* genomes	Illumina HiSeq 2000	20 genomes: >50x 25 genomes: <36x	1.04 – 1.23	SOAPdenovo (Lou *et al.*, 2012).	10 - 34	0.3 – 10.6

Bp: base pairs
1 Kbp = 1 000 Bp
1 Mbp = 1 000 000 Bp
Gb = 1,000,000,000 bp.
x: Coverage of the genome as times.
* CLC genomics reference: https://www.qiagenbioinformatics.com/?result=citations&s=

From Table 3.1 can be concluded that most of the avian genomes sequenced here were in the size range between 1.04 – 1.2 Gbp. The Puerto Rican parrot genome was however found to be slightly larger (1.58 Gbp). Since the budgerigar is the closest relative of the lovebird with a sequenced genome it can be estimated that the lovebird genome will range between 1.0 Gb to 1.2 Gb in size.

Sims *et al.* (2014) defines expected or theoretical genome coverage during sequencing as the average number of times a nucleotide is expected to be sequenced given a specific number of reads, all of approximately the same length and randomly dispersed throughout the genome. This coverage (in times or x) influences the accuracy of the assembly. Refer to Table

3.1 for the coverage of all the species' genomes. Sequencing the genome at a higher coverage will have a direct positive effect on metrics like the scaffold- and contig-N50 lengths, (metrics most commonly used to assess the completeness and accuracy of an assembly). N50 lengths are calculated by adding all reads' lengths from the longest until half (50%) of the total assembly length is reached (Baker, 2012; Bradnam *et al.*, 2013). All genomes in the Zhang-study sequenced at a high (>50x) coverage had N50 scaffold lengths in excess of 1 Mbp except the white-throated Tinamou (*Tinamous guttatus*) (242 Kbp) and the bald eagle (*Haliaeetus leucocephalus*) (670 Kbp). This was due to the absence of the 12 Kbp and 20 Kbp libraries for these two genomes. There is no "golden number" when it comes to N50 lengths and generally "longer is better" is the only guideline. Additional sequencing data, either libraries with larger insert sizes (e.g. 12 Kbp or 20 Kbp) or sequencing platforms (e.g. combining different platform data e.g. Roche-454 or Illumina data with PacBio data) could improve these metrics. PacBio sequencing platform yield longer read lengths than any other platform, but the accuracy is lower (Quail *et al.,* 2012). It should however be noted that PacBio analyses are far more expensive than any other platform. However, in the budgerigar study the addition of longer reads did not have a profound effect on the scaffold N50 lengths, but the coverage did (refer to Table 3.1). The genomes sequenced at a higher coverage (>50x) during the Zhang study all yielded higher N50 lengths. If a genome can be sequenced at a higher coverage on the Illumina or Roche platforms it might have a more profound effect on the N50 scaffold lengths, but a genome might be assembled more complete (less gaps between scaffolds) when using longer reads (PacBio). Therefore, the aim of the study should be kept in mind when deciding on sequencing coverage, the type of platform and the number of sequencing libraries to use. If the aim is to assemble a complete genome, then longer reads could be beneficial to this process. If the aim is to identify SNPs throughout the genome and the location isn't of great concern, then a first draft genome should be sufficient. The aim of the *Agapornis* study was to develop genetic screening tests, such as a SNP-based parentage verification test. Therefore, the completeness of the genome was not the main focus of this study.

In an attempt to address the issue of inconsistent genome assemblies as well as trying to objectively find assembly tool benchmarks, the Assemblathon and Assemblathon 2 projects were launched (Earl *et al.*, 2011; Bradnam *et al.,* 2013). During these projects sequence data from different organisms were sent to different teams to evaluate their assemblies of the same data. During Assemblathon 2, the budgerigar was one of the three organisms that was assembled (Bradnam *et al.*, 2013) using the same data as described in Ganapathy *et al.* (2014). Twenty-one teams from across the globe received the data and used different assemblers to assemble the genomes. The results from different teams displayed great

variation in the genome size and contig and scaffold lengths, even though the same dataset was used (Bradnam *et al*., 2013). Bradnam *et al.* (2013), after reviewing the Assemblathon 2 assemblies, suggest that a single metric (e.g. N50) should not be used to assess the quality of an assembly and the aim of the study should be kept in mind when it is assessed. N50 values don't give an indication of e.g. gene content (Baker, 2012; Bradnam *et* al., 2013; Simão *et al.*, 2015) therefore other metrics in addition to N50 length such as the number of scaffolds that are gene-sized, the number of core genes that are present, the number of gaps in the assembly as well as scaffold and contig statistics should be analysed too (Bradnam *et al*., 2013; Wajid & Serpedin, 2014). One metric used to assess the Assemblathon 2 genomes was that of the presence of highly conserved genes that should be present in nearly all eukaryotic genomes, or core eukaryotic genes (CEG). Only one of the assemblies contained more than 95% of the CEG, whereas the majority contained between 85-90% of these genes. The assembly with the highest CEG's also had the highest scaffold N50 lengths. They made use of two assemblers, ALLPATHS-LG and Velvet. The use of additional assemblers could therefore also be beneficial to improve on the assembly statistics.

In summary, these studies showed that by sequencing a genome at a coverage >50x (high coverage) and using an assembler, such as SOAPdenovo, ALLPATHS or Velvet, designed for *de novo* genome assembly, it should generate an assembly that is both accurate and complete. The use of a specific sequencing platform did not have a major effect on the genome assembly, but by using PacBio reads the number of gaps could be reduced. A single metric, such as N50 lengths, should not be viewed as the only metric to assess the assembly. Additional metrics as described above, especially the presence of CEG, should be applied too.

Paper II: Draft de novo genome Sequence of Agapornis roseicollis for Application in Avian Breeding

Henriëtte van der Zwan, Francois van der Westhuizen, Carina Visser and Rencia van der Sluis.

Published in: Animal Biotechnology (2018) 29:4, 241-246.

Taylor & Francis

Animal Biotechnology

ISSN: 1049-5398 (Print) 1532-2378 (Online) Journal homepage: http://www.tandfonline.com/loi/labt20

Draft De Novo Genome Sequence of *Agapornis roseicollis* for Application in Avian Breeding

Henriëtte van der Zwan, Francois van der Westhuizen, Carina Visser & Rencia van der Sluis

To cite this article: Henriëtte van der Zwan, Francois van der Westhuizen, Carina Visser & Rencia van der Sluis (2018) Draft De Novo Genome Sequence of *Agapornis roseicollis* for Application in Avian Breeding, Animal Biotechnology, 29:4, 241-246, DOI: 10.1080/10495398.2017.1367692

To link to this article: https://doi.org/10.1080/10495398.2017.1367692

Published online: 16 Oct 2017.

ANIMAL BIOTECHNOLOGY
2018, VOL. 29, NO. 4, 241–246
https://doi.org/10.1080/10495398.2017.1367692

Draft De Novo Genome Sequence of *Agapornis roseicollis* for Application in Avian Breeding

Henriëtte van der Zwan[a], Francois van der Westhuizen[a], Carina Visser[b] and Rencia van der Sluis[a]

[a]Centre for Human Metabolomics, North-West University, Potchefstroom, North-West, South Africa; [b]Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, Gauteng, South Africa

ABSTRACT

In aviculture, lovebirds are considered one of the most popular birds to keep. This African parakeet is known for its range of plumage colors and ease to tame. Plumage variation is the most important price-determining trait of these birds, and also the main selection criterion for breeders. Currently, no genetic screening tests for traits of economic importance or to confirm pedigree data are available for any of the nine lovebird species. As a starting point to develop these tests, the de novo genome of *Agapornis roseicollis* (rosy-faced lovebird) was sequenced, assembled, and annotated. Sequencing was done on the Illumina HiSeq 2000 platform and the assembly was performed using SOAPdenovo v2.04. The genome was found to be 1.1 Gb in size and 16,044 genes were identified and annotated. This compared well with other previously sequenced avian genomes, such as the chicken, zebra finch, and budgerigar. To assess genome completeness, the number of benchmarking universal single-copy orthologs were identified in the genome. This was compared to other previously assembled avian genomes and the results indicated that the genome will be useful in the development of genetic screening tests to aid lovebird breeders in selecting breeding pairs.

KEYWORDS

Avian genomics
bioinformatics; whole genome sequencing

Introduction

Birds from the genus *Agapornis* (commonly called lovebirds) form part of the subfamily Psittacinae, or parrots. There are 356 parrot species distributed over Africa, Asia, Australia, and South America and are recognized by their blunt bill with a downward curving upper mandible (1). There are nine lovebird species, eight of which are native to Africa and one to Madagascar (2). *Agapornis roseicollis* (rosy-faced lovebird) is indigenous to the South-Western parts of Africa including Angola, Namibia, and South-Africa (2, 3). Wild flocks are still found, however, lovebirds are best known as pets due to the fact that they are easily tamed and breed effortlessly in captivity (3).

There are 20 naturally occurring *A. roseicollis* color variations, all of which are inherited as Mendelian traits, either autosomal or sex-linked recessive or dominant. Some of these variants are shown in Fig. 1. The plumage color dictates the price of the bird so breeders experiment by crossing birds of different plumage colors attempting to breed birds with unique plumage colorations. Some of the species can also be interbred to produce viable hybrid offspring, often with spectacular color variations (Personal communication, Mr D van den Abeele).

Parrots have a unique mechanism of coloration and unlike most bird species do not utilize carotenoids to generate red, yellow, orange, green, blue, or violet plumage colors (4–8). Parrot pigments differ both chromotographically and spectrally from all other feather carotenoids (6, 8) and is called psittacofulvins. Psittacinae is the only aves family that express psittacofulvins in their plumage cells. Although the physiological or anatomical sources of these molecules are still unknown, it is strongly believed that diet does not play a role in parrot coloration.

Due to the strong economic influence, the main trait breeders are selecting for is color. As most of the colors are inherited in a Mendelian recessive manner, chicks born from parents with color variants, but that display wild type coloration, may be heterozygous for the rare color. Since there is no screening test available, bird sellers cannot guarantee that the chick is a heterozygote of that color, nor can they confirm the parentage of the chick. Developing a Single Nucleotide Polymorphism (SNP)-based test to genotype individuals could also

CONTACT Henriëtte van der Zwan henriettevdz@gmail.com Centre for Human Metabolomics, North-West University, Hoffman Street, Potchefstroom, North-West 2531, South Africa.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/labt.

Figure 1. Some of the plumage variations of *Agapornis roseicollis*: (a) wild type, (b) opiline-D blue, (c) orange mask and (d) aqua. (Photos curtesy of D van den Abeele).

result in a reduction of illegal parrot trade. It is estimated that 300,000 illegal parrots are sold in the United States every year. This can be lowered if genotyping of birds can be done on a routine basis (9).

In an attempt to address the lack of genetic screening tests for lovebird breeders worldwide, the first step was to perform a de novo genome sequencing and assembly of *A. roseicollis*. Three parrot genomes namely the Puerto Rican parrot (*Amazona vittata*) (10) scarlet macaw (*Ara macao*) (11), and the budgerigar (*Melopsittacus undulates*) (12) have been sequenced previously. In 2014, Zhang et al. published the genomes of 48 bird species including that of the budgerigar (13).

Here we describe the first de novo attempt to sequence, assemble and annotate the genome of *A. roseicollis*.

Materials and methods

Sample selection

One adult male (1 year in age), that was born and bred in captivity in Belgium, was selected as the *A. roseicollis* reference genome. The criteria set for sample selection was the availability of a minimum of five generations pedigree data as well as information on all plumage colors of its ancestors. After ethical approval was obtained from the North-West University AnimCare committee (Ethics number NWU00348-15-S5), 400 µL blood collected by a veterinarian in an EDTA tube was taken from the bird. Blood samples were shipped on ice to Germany for sequencing.

Whole genome sequencing

Genomic DNA was isolated from the blood sample using the Machery Nagel Blood Mini kit according to manufacturer's protocol with one change in that 10 µL blood diluted with 190 µL Phosphate-buffered saline (PBS) were used as starting material. Sequencing was performed by Eurofins Genomics in Germany. The sequencing was done at an average total depth of 100 × coverage.

Three shotgun libraries were constructed consisting of fragment sizes of 300, 550, and 750 kb, respectively. Creation of the Shotgun library was done using commercially available kits (NEBNext® Ultra™ DNA Library Prep Kit for Illumina, article number E7370) according to the manufacturer's instructions. In total, 1 µg of DNA was fragmented using a Covaris E210 Instrument (Covaris Inc., Woburn, MA) according to manufacturer's instructions. End-repair, A-tailing, and ligation of indexed Illumina Adapter, size selection, and amplification was performed accordingly. The resulting fragments were cleaned up and pooled. In addition, one 3 kb and one 8 kb mate-pair-like library (LJD library, Eurofins Medigenomix, Ebersberg, Germany, proprietary protocol) were prepared based on the mate-pair library protocol from Illumina, modified using adaptor-guided ligation of genomic fragments which achieves higher accuracies. The resulting fragments were cleaned up and pooled.

The libraries were analyzed on the Illumina HiSeq 2000 platform with chemistry version 4 and sequencing was performed using manufacturer's instructions. Paired-end sequencing using 125 bp read length was performed on a HiSeq machine using v4 chemistry (HiSeq Control Software 2.2.38). For processing of raw data RTA version 1.18.61 and CASAVA 1.8.4 was used to generate FASTQ-files.

Genome assembly

The lovebird genome assembly and annotation was performed at BGI in collaboration with the Bird 10 K (B10 K) genome project (14). The lovebird genome was assembled using de novo software SOAPdenovo v2.04 (15). Filtering of low quality reads were done on the basis of *k*-mer frequency error correction by removing reads with more than 10% ambiguous bases. Reads

were filtered and removed if more than 40% of bases were of low quality and duplicate reads were removed.

De Bruijn graphs were constructed by SOAPdenovo (15) and then tips were clipped, bubbles merged, and low coverage links removed. Contigs were collected using *k*-mer lengths of 69 and all usable libraries reads were mapped to contig sequences to construct scaffolds. Pair-end reads were used to fill gaps between scaffolds as one end of the read will align with a contig and the other end with the gap. The Gapcloser module of SOAPdenovo (15) was used to fill gaps within scaffolds. The genome was uploaded onto NCBI with accession number NDXB01000000.

Genome annotation

Transposable elements were identified by executing RepeatProteinMask in RepeatMasker 4.0.5 (16) by comparing the assembly against the Repbase TE library (Repbase-17.06). This identifies transposable elements by aligning the genome sequence to a self-taken curated transposable element protein database. The default parameters of RepeatModeler (16) was used to build a de novo lovebird repeat library and this resulted in consensus sequences and classification information for each repeat gene family.

Gene annotation was done in two parts—homology-based gene prediction and gene function annotation. During the homology-based gene prediction stage, Ensembl gene sets (release 60) (17) of the chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and human (*Homo sapiens*) genomes were used to annotate the protein-coding genes. These species were selected based on the fact that their genomes were sequenced with a very high coverage and the annotation of genes will more likely be correct. If future gene studies are to be conducted on this genome it is advisable to use closer related species, e.g., the budgerigar, genomes to reannotate the genome. This gene set's protein sequences were used as reference templates for homology-based gene predictions. The gene function annotation phase relied on the annotation of the motifs and domains of the reference genes using InterPro. Databases in the public domain were used for this part of the annotation. Gene products were presented by Gene Ontology and retrieved from InterPro results. Finally, the reference genes were mapped to Swiss-prot database to find the best match for each gene (18).

The annotation pipeline was executed by first performing a rough alignment where protein sequences of the reference gene set was aligned using TBLASTN. The cut-off value was $1e^{-5}$. genBlastA was used to find the corresponding gene loci with the blast hits. All loci with homologous block lengths shorter than 30% of the query protein were excluded. A precise alignment followed where sequences of candidate gene loci were extracted and using GeneWise v2.2.0, the precise alignment was concluded. MUSCLE v3.8.31 was then run on the predicted protein and reference protein, where-after predicted proteins were filtered out if shorter than 30 aa or had a percent identity less than 25% as well as pseudogenes. Lastly a non-redundant gene set was built to ensure that no gene overlaps were annotated.

Comparing avian genomes

The lovebird genome was compared to other avian genomes to assess its size, number of genes, scaffold N50 size, contig N50 size, and gene completeness. The chicken (Gallus_gallus-5.0) (*G. gallus*) and zebra finch (*T. guttata*) genomes were selected as they are considered as the avian model organisms. The budgerigar (*M. undulates),* kea (*Nestor notabilis*), Peregrine falcon (Falco peregrinus), Puerto Rican parrot (*A. vittata*), and Scarlet macaw (*A. macao*) were included as they are close relatives of the lovebird.

Assessing gene completeness

Bradnam et al. (21) report the metrics used to assess the quality of a genome. In addition to Scaffold N50 length, Contig N50 length and the number of scaffold sequences that are gene sized, the number of core genes mapped in the annotation is also important. Making use of the open-source software that is implemented in Python, benchmarking universal single-copy orthologs (BUSCO) v2 (22) the lovebird genome as well as seven other previously assembled bird genomes were assessed based on gene completeness using the above mentioned method. All genomes were compared to a set of 303 conserved eukaryote genes namely OrthoDB v9.1. OrthoDB v9.1 is a set of orthologs that is used by BUSCO v2 to assess gene completeness (23). OrthoDB covers 5,756 species including bacteria, eukaryotes, fungi, plants, archaea, and viruses.

Results and discussion

The lovebird genome was estimated to consist of 1,159,816,267 base pairs. Sequencing coverage of each sequencing library can be calculated by dividing the number of bases generated by the size of the genome. The sequencing coverage per library is shown in Table 1. The contig N50 and scaffold N50 lengths were 5,455 and 108,514, respectively. The G/C content of the genome was 43%.

Table 1. Sequencing coverage per library.

Sample and library	Yield (Mbp)	Size of genome	Sequencing coverage per library
Offspring_300	33,970	1,159,816,267	58.02
Offspring_550	14,995	1,159,816,267	25.62
Offspring_750	11,824	1,159,816,267	20.30
Offspring_3kb	29,934	1,159,816,267	25.81
Offspring_8kb	13,426	1,159,816,267	11.58

Zhang et al. (13) reported the assembled genomes of 48 bird species. From these species, five species were selected for this study and the genomes compared. Two additional parrot species, the Puerto Rican parrot (10) and Scarlet macaw (11), were also included as shown in Table 2.

The size of the lovebird genome was comparable to all seven of the other genomes with the exception of the Puerto Rican parrot. The larger size of the Puerto Rican parrot could be due to a poor assembly and lower genome coverage because it is larger than any of the avian genomes sequenced by Zhang et al. (13). The lovebird genome was sequenced at 100× coverage and the scaffold N50 lengths were shorter than comparative genomes. This could be as a result of shorter read lengths resulting in shorter contigs and scaffolds. Zhang et al. (13) also note that the use of 20 kb insert libraries will increase these metrics and it was not included in this study. Since scaffold N50 length does not give an indication on gene content, and in this study gene function is of greater importance than genome completeness, the lower N50 scaffold and contig values are not of great concern (21, 22).

During the genome annotation, 15,045 gene coding sequences were identified and 999 noncoding sequences were identified. This compares well with the other genomes, with the exception of the chicken genome, where between 18,618 (zebra finch) and 14,074 (kea) genes were annotated (Table 2). The chicken genome (v Gallus_gallus-5.0) was found to have 26,640 genes (19). It should, however, be noted that this is the fifth draft of the chicken genome and that it has been sequenced using Sanger sequencing as well as various other next generation sequencing platforms and, therefore, is more complete and accurate.

By comparing the genome to a eukaryotic dataset comprising of 303 genes from OrthoDB v9.1 (23) we found 258 or 85.2% complete BUSCOs. Simão et al. (22) defines a complete BUSCO as a gene with a length within two standard deviations of the BUSCO gene group mean length. About 7.3% of the genes were fragmented which is defined as genes only partially recovered and 7.5% of genes were missing indicating that they were not identified at all, as displayed in Fig. 2.

Benchmarking universal single-copy orthologs were also analyzed in the other seven avian genomes that were compared and the results shown in Table 3.

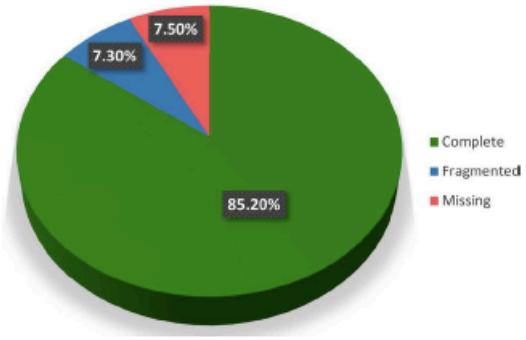


Figure 2. Genome completeness assessment of *Agapornis roseicollis* using BUSCO. *Note*: BUSCO, benchmarking universal single-copy ortholog.

Table 3. Comparison of BUSCOs between different avian genomes.

Organism	Complete BUSCOs (%)	Missing BUSCOs (%)	Fragmented BUSCOs (%)
Agapornis roseicollis (Lovebird)	85.2	7.5	7.3
Melopsittacus undulates (Budgerigar)	81.2	14.2	4.6
Gallus gallus (Chicken)	88.8	10.5	0.7
Taeniopygia guttata (Zebra finch)	81.9	5.6	12.5
Amazona vittata (Puerto Rican parrot)	59.5	16.1	24.4
Ara macao (Scarlet macaw)	45.2	30	24.8
Falco peregrinus (Peregrine falcon)	88.2	7.5	4.3
Nestor notabilis (Kea)	70.3	14.8	14.9

BUSCOs, benchmarking universal single-copy orthologs.

Table 2. Comparison of avian genomes.

Organism	Genome size (bp)	Scaffold N50	Contig N50	Number of genes	Sequence depth and platform	Reference
Agapornis roseicollis (Lovebird)	1.1 G	109 K	5.4 K	16,044	100×, Illumina	Current study
Melopsittacus undulatus (Budgerigar)	1.2 G	10.6 M	55.6 K	16,240	160×, Multiple platforms	(12)
Gallus gallus (Chicken) v Gallus_gallus-5.0	1.21 G	6.3 M	2.8 M	26,640	50.6×, PacBio RSII	(19)
Taeniopygia guttata (Zebra finch)	1.2 G	10.4 M	38.5 K	18,618	6×, Sanger	(20)
Amazona vittata (Puerto Rican parrot)	1.58 G	19.47 K	6.9 K	N/A	27×, Illumina	(10)
Ara macao (Scarlet macaw)	1.2 G	15.9 K	6.3 K	14,405	16×, Roche 454; 13×, Illumina	(11)
Falco peregrinus (Peregrine falcon)	1.2 G	3.89 M	28.5 K	16,263	105×, Illumina	(13)
Nestor notabilis (Kea)	1.14 G	37 K	16 K	14,074	32x, Illumina	(13)

The number of complete BUSCOs identified from the lovebird genome corresponds well with the number of complete BUSCOs identified in other genomes, such as the budgerigar, chicken, and zebra finch. The lower number of complete BUSCOs identified in the kea, Puerto Rican parrot, and Scarlet macaw genomes could be due to lower genome coverage during sequencing.

Conclusion

In conclusion, this study and results provides the first genomic information in the form of a full genome sequence, assembly, and comparison of the genus *Agapornis*. The results indicate that the lovebird genome contains an adequate number of core genes to be useful in further research such as the identification of SNPs for parentage verification, investigating the genetic basis linked to color variation, and the development of tests to distinguish hybrids of different species.

Acknowledgments

We want to thank Mr Dirk van den Abeele for his help on selection of the samples. Without the help of the Avian Phylogenomics Project, in particular Professor Guojie Zhang, Professor Tom Gilbert and Professor Erich Jarvis as well as Dr Cai Li and Dr Yuan Deng from BGI, the assembly and annotation would not have been possible.

Funding

This work was supported by the Technology and Human Resources for Industry Programme (THRIP) which forms part of the Department of Trade and Industry of the South African government (project ID number TP13082831254) and Lumegen Laboratories.

References

1. Forshaw JM. *Parrots of the World*. United States of America: Princeton University Press; 2010.
2. Dilger WC. The comparative ethology of the African parrot genus *Agapornis*. *Ethology* 1960; 17(6):649–685. doi:10.1111/j.1439-0310.1960.tb00412.x
3. Van den Abeele D. Lovebird Compendium. *The Netherlands: About Pets/Welzo Media Productions BV*; 2016.
4. Dyck J. Structure and spectral reflectance of green and blue feathers of rose-faced lovebirds (*Agapornis roseicollis*). *Biol Skr Dan Vid Selsk* 1971; 18:1–67.
5. Dyck J. Structure and colour-production of the blue barbs of *Agapornis roseicollis* and *Cotinga maynana*. *Zeitschr Zellforsch* 1971; 115:17–29. doi:10.1007/bf00330211
6. Hudon J, Brush AH. Identification of carotenoid pigments in birds. *Methods Enzymol* 1992; 213:312–321. doi:10.1016/0076-6879(92)13133-i
7. McGraw KJ, Nogare MC. Carotenoid pigments and the selectivity of psittacofulvin-based coloration systems in parrots. *Comp Biochem Physiol B* 2004; 138:229–233. doi:10.1016/j.cbpc.2004.03.011
8. Hill GE, McGraw KL. (eds). *Bird Coloration Mechanism and measurements*. Vol. 1. United States of America: Harvard University Press; 2006.
9. Pires S. The illegal parrot trade: a literature review. *Global Crime* 2012; 13:3.
10. Oleksyk TK, Pombert J-F, Siu D, Mazo-Vargas A, Ramos B, Guiblet W, Afanador Y, Ruiz-Rodriguez CT, Nickerson ML, Logue DM, Dean M, Figueroa L, Valentin R, Martinez-Cruzado J-C. A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. *GigaScience* 2012; 1:14.
11. Seabury CM, Dowd SE, Seabury PM, Raudsepp T, Brightsmith DJ, Liboriussen P, Halley Y, Fisher CA, Owens E, Viswanathan G. Tizard, IR multi-platform draft de novo genome assembly and comparative analysis for the Scarlet Macaw (*Ara macao*). *PLoS ONE* 2012; 8:e62415. doi:10.1371/journal.pone.0062415
12. Ganapathy G, Howard JT, Ward MJ, Li J, Li B, Li Y, Xiong Y, Zhang Y, Zhou S, Schwartz DC, Schatz M, Aboukhalil R, Fedrigo O, Bukovnik L, Wang T, Wray G, Rasolonjatova I, Winer R, Knight JR, Koren S, Warren WC, Zhang G, Phillippy AM, Jarvis ED. High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience* 2014; 3:11.
13. Zhang G, Li B, Gilbert MT, Jarvis ED, Wang J, The Avian Genome Consortium. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 2014; 346:6215.
14. Zhang G. Bird sequencing project takes off. *Nature* 2015; 522:34.
15. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 2012; 1:18.
16. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. http://www.repeatmasker.org. 2013.
17. Flicek P, Ahmed I, Amode R, Barrell Barrell. Ensembl. 2013. *Nucleic Acids Res* 2012; 1:8.
18. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acid Res* 2015; 43:D204–D212.
19. Warren WC, Hillier LW, Tomlinson C, Minx P, Kremitzki M, Graves T, Markovic C, Bouk N, Pruitt KD, Thibaud-Nissen F, Schneider V, Mansour TA, Brown CT, Zimin A, Hawken R, Abrahamsen M, Pyrkosz AB, Morisson M, Fillon V, Vignal A, Chow W, Howe K, Fulton JE, Miller MM, Lovell P, Mello CV, Wirthlin M, Mason AS, Kuo R, Burt DW, Dodgson JB, Cheng HH. A new chicken genome assembly provides insight into Avian genome structure. *G3* 2017; 7:109–117.
20. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TAF, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin Y-C, George J, Sweedler

J, Southey B, Gunaratne P, Watson M. The genome of a songbird. *Nature* 2010; 464:757–762.

21. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C, Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G, Gibbs RA, Gnerre S, Godzaridis E, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S, Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li Y, Li Z, Liu B, Liu Y, Luo R, MacCallum I, MacManes MD, Maillet N, Melnikov S, Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG, Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J, Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2013; 2:10.
22. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015; 31:19.
23. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res* 2015; 45:D744–D749.

CHAPTER FOUR: DEVELOPMENT OF A SNP-BASED PARENTAGE VERIFICATION PANEL FOR LOVEBIRDS.

4.1 INTRODUCTION

Discovering SNPs from non-model organisms can be challenging, especially if a reference genome is lacking. As described in section 2.7, discovering variants using whole genome resequencing (WGR), once a reference genome is available, is an economical and highly efficient method to apply. WGR entails sequencing the whole genome of an individual at a lower coverage and by comparing those reads to the reference genome, differences in the nucleotides (possible SNPs) are identified. These SNPs can then be filtered to include the best callset to be used for the application needed such as to identify or verify parentage. One of the programs available to use as a variant caller in WGR is Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). GATK is executed in the command-line interface and gives clear set best-practices guidelines. In this chapter more detail on the GATK pipeline is described.

In this study, the genomes of the two parents of the reference genome individual were used to identify SNPs in order to evaluate Mendelian inheritance of SNPs. In Figure 4.1 the two parents of the reference genome chick can be seen.

Figure 4.1: The parents of the reference genome chick.
Photo: Henriëtte van der Zwan

4.2. AN IN DEPTH DISCUSSION OF THE GATK PIPELINE

The GATK pipeline is divided into three stages – pre-processing, variant discovery and callset refinement. Callset refinement was not executed in this study. An overview of this process is given in Figure 4.2.

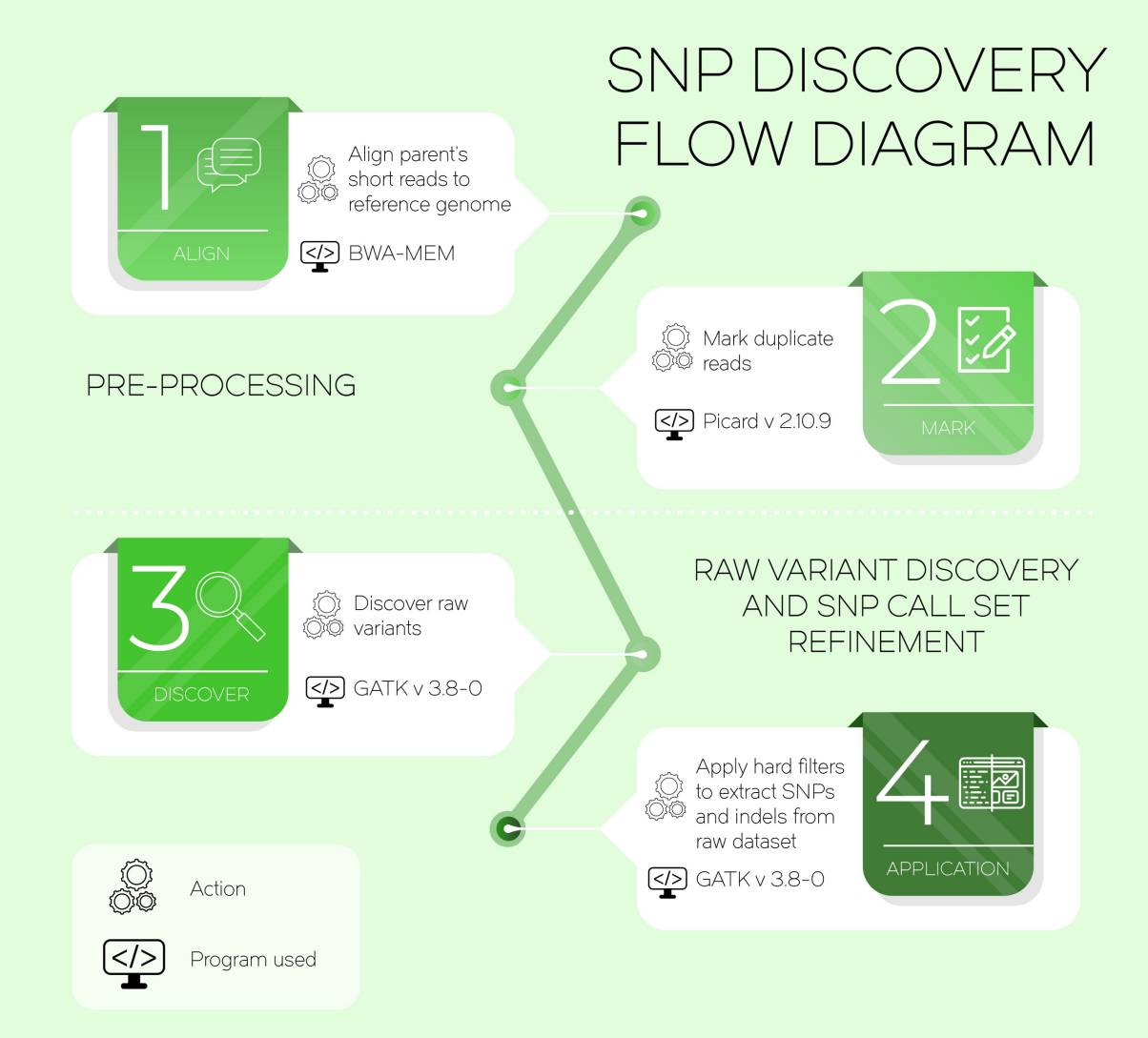


Figure 4.2: Best practises guidelines followed, as set out by GATK. Graphic: H. van der Zwan

The pre-processing steps are not executed in GATK and only serves to prepare the data to be used in GATK. This entails mapping the short reads (in this case the WGR reads of both parents) to the reference genome using BWA-mem (Li & Durbin, 2009). The second step entails the marking of duplicate reads using Picard (http://broadinstitute.github.io/picard/). This reduces the number of erroneous bases called during sequencing. Variant discovery is executed in GATK using the GVCF (Genomic Variant Call Format) module and HaplotypeCaller function. The output is a single set of raw variants that includes SNPs, indels (insertions or deletions), mixed variants (combinations of SNPs and indels at a single position), MNP (multi-nucleotide polymorphisms e.g. dinucleotide substitution) and symbolic variants (e.g. the * symbol indicate a spanning deletion). Genotypes of all samples can be combined in this step to make SNP selection easier as SNPs from all samples are combined.

To filter out false positive variants as well as all unwanted variants (e.g. indels) GATK's Variant Quality Score Recalibration (VQSR) can be used for model organisms. For non-model organisms (such as the lovebird) hard filtering has to be applied. Hard filtering is done in two stages by firstly extracting the true SNPs from the call-set and then extracting the indels from the call-set. The parameters used to filter the SNPs and indels include QualByDepth (QD), FisherStrand (FS), RMSMappingQuality (MQ), MappingQualityRankSumTest (MQRankSum) and ReadPosRankSumTest (ReadPosRankSum) (https://software.broadinstitute.org/gatk/img/BP_workflow_3.6.png and De Summa *et al*., 2017). In Table 4.1 these parameters and an explanation of each is given.

Table 4.1: Parameters recommended to use by GATK during hard filtering of SNPs and indels.

Parameter	Details
QUAL	A Phred-based probability of a false positive variant.
QD	QD is computed by dividing the variant confidence by the unfiltered depth of non-homozygous reference samples. Most variants lay within a QD value of 10 – 15. Homozygous samples contribute double the number of reads than heterozygous samples, therefore the majority of variants around those values.
FS	FS is a Phred-scaled probability that a site is bias towards a strand, indicating that the alternative allele is more or less likely to be seen on a specific strand. A value of 0 indicates that there is little to no strand bias.
MQ	MQ is calculated as the root mean square mapping quality of all the reads at that site and include the standard deviation of the mapping qualities and thereby including the variation in the dataset. Values around 60 indicate a good mapping quality.
MQRankSum	MQRankSum compares the mapping qualities of the reads of the reference vs the alternative allele. A positive MQRankSum value indicates that the quality of the alternative allele is higher than the reference allele where a negative value indicates the opposite. A value of zero indicates no difference in the mapping qualities of the two alleles.

ReadPosRankSum	ReadPosRankSum defines if the position of the alternative and reference alleles is the same within the reads. A negative value indicates that the alternative allele is found at the end of a read more often, a positive value that the reference allele is found at the end more often: and a value of zero that there is no significant difference in the positions.

After hard filtering the output file is a set of SNPs and a set of indels matching the filters. The third and final stage of the GATK analysis is to refine the set of variants. This step was not applied in this study since only two individuals were used to identify SNPs and annotation was redundant for a parentage verification panel. GATK's recommendations, should this step be followed, is that the genotypes in the dataset are redefined by using additional information, such as pedigree data. This data helps to improve the accuracy of the genotypes of the SNPs. Afterwards, the variants are annotated based on context and function. Context annotation can be done using the VariantAnnotater module of GATK but there is no functional annotation module available on GATK. In Appendix E, Supplementary File 1, the command line arguments followed during the GATK analysis in this study, is shown.

4.3 SUBMITTED MANUSCRIPT I

A full length manuscript of the WGR, SNP identification and parentage verification panel construction is given here. Supplementary files of this manuscript is given in Appendix E. A short communication manuscript was submitted to Animal Genetics (Manuscript ID AnGen-19-05-0148) (Submitted Manuscript I) and shown in Appendix F.

Development of a SNP-based parentage verification panel for lovebirds

Henriëtte van der Zwan*[1], Carina Visser[2], Maryke Schoonen[1] & Rencia van der Sluis[1]
[1]Focus Area for Human Metabolomics, North-West University, Potchefstroom, South Africa
[2]Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa
*Corresponding author Email: agapornis.genome.study@gmail.com

Abstract

The genus *Agapornis*, or lovebirds, are popular pet parrots worldwide. Breeders depend on pedigree records as a selection tool as no molecular parentage verification test is available for any of the nine species. The reference genome of *A. roseicollis* was recently assembled. The whole genome of the parents of the reference genome individual were sequenced at 30x coverage to identify single nucleotide polymorphisms (SNPs) by mapping the reads against the reference genome. By applying the Genome Analysis Toolkit (GATK) pipeline, over 1.6 million SNPs, shared between the parents' genomes, were discovered. SNPs were filtered to a panel of 480 SNPs based on GATK parameters. This panel was genotyped in a population of 960 lovebirds across seven species. The Father's sequence was used as the reference for the genotyping results and SNPs that did not amplify in this sample were discarded, resulting in a panel of 262 SNPs. This panel was reduced based on the Observed Heterozygosity (H_O) and Minor Allele Frequency (MAF) values per SNP to include the lowest number of SNPs with the highest exclusion power for parentage verification. Two additional panels consisting of 195 SNPs with MAF and H_O values >0.1 and 40 SNPs, where these values were >0.3, were constructed. Parentage verification was performed on 43 families from different species with known relationships to assess the exclusion power of each panel. The 195-SNP panel with an average exclusion probability of 99.9% and MAF and H_O values >0.1 was proposed as the *Agapornis* parentage verification panel.

Keywords: *parrot breeding, pedigree confirmation, whole genome resequencing, Agapornis*

Introduction

The parrot genus *Agapornis* (or lovebirds) consists of nine African parrot species that are commonly kept as companion animals (Dilger, 1960; Van den Abeele, 2016). They are also found in their natural habitat across Africa and Madagascar (Forshaw, 1989). Eight of the species are found in captive breeding populations (*A. roseicollis*, *A. fischeri*, *A. lilianae*, *A. nigrigenis*, *A. personatus*, *A. canus, A. pullarius* and *A. taranta*). Lovebird breeding is popular across the globe due to these birds being hardy and active, they also breed easily in captivity,

do not need big aviaries and display a range of colour variations (Hayward, 1979). Eberhard (1998) and Manegold and Podsiadlowski (2014) sequenced a portion of the cytochrome b (cytb) mitochondrial gene of each of the nine *Agapornis* species. They found *A. fischeri*, *A. personatus*, *A. lilianae* and *A. nigrigenis* to group together (called the white eye-ring group due to the patch of white skin around the eye), *A. taranta*, *A. canus* and *A. pullarius* (the latter was not included in this study) grouped together (called the sexually dimorphic group due to males and females displaying different plumage coloration) and *A. roseicollis* as a transitional group (no white ring is present but males and females are identical in plumage coloration). *A. swindernianus* (not included in this study) was found to be a sister clade of all other species. In aviculture *A. fischeri*, *A. personatus* and *A. roseicollis* were voted in a recent poll as the most popular breeding species (https://www.ornitho-genetics.info/), *A. lilianae* and *A. nigrigenis* are found less frequently in breeding systems whereas *A. canus* and *A. taranta* are unpopular. *A. pullarius* breeds poorly in captivity and is, therefore, very seldom seen in breeding systems while *A. swindernianus* doesn't survive in captivity (Dilger, 1960; Van den Abeele, 2016).

Plumage colour is the main attraction for breeders and buyers, and is therefore the main selection criterion used by breeders. Most of the 30 observed colour variations are inherited as autosomal or sex-linked recessive traits but the causative mutations associated with these variations have not yet been identified (Hayward, 1979; Van den Abeele, 2016; van der Zwan *et al.*, 2019). Due to the inheritance patterns of these traits, breeders make use of pedigree data to predict the heterozygous genotype for a specific colour. Despite the dependency on pedigree data, no species-specific parentage verification test is available for *Agapornis*. As a result, many fraudulent transactions take place as sellers know there is no molecular test available to verify either the pedigree or the colour genotype of a chick. Accurate and complete pedigree records are of utmost importance in any animal or avian breeding system, but even more so where pedigrees are used to select offspring. Currently, registration bodies do not require molecular parentage verification to register, import or export birds.

Two recent studies developed microsatellite-based parentage verification panels for parrot species. Coetzer *et al.* (2017) developed a panel of 16 markers with sufficient power to exclude non-parents for the Cape parrot (*Poicephalus robustus*). Jan & Fumagalli (2016) identified 106 microsatellite markers which amplified across seven different parrot species (*Amazona brasiliensis*, *A. oratrix*, *A. pretrei*, *A. rhodocorytha*, *Anodorhynchus leari*, *Ara rubrogenys* and *Primolius couloni*) for application of individual identification. Despite the effectivity of microsatellites to verify parentage in many domestic animal species, the use of microsatellite markers has generally become redundant and it has been replaced with Single Nucleotide Polymorphisms (SNPs). The main reasons are the inconsistencies in allele size calling of

microsatellite loci making data sharing between laboratories difficult (Vignal *et al.*, 2002; Garvin *et al*., 2010) and genotyping errors due to false "new" alleles being incorrectly called. Due to the bi-allelic nature of SNPs an individual can only be homozygous for the one or the other nucleotide or a heterozygote, (Vignal *et al*., 2002). In scenarios where an individual was previously genotyped (often at another laboratory) and this profile is necessary to confirm kinship, accurate data exchange is of utmost importance. The bi-allelic nature of SNPs results in a lower exclusion power than multi-locus microsatellite markers and therefore more SNPs are needed compared to microsatellite markers (Olsen *et al.*, 2011).

Three previous avian SNP-based parentage verification panels have been constructed, namely for African Penguins (*Spheniscus demersus*) (Labuschagne *et al., 2015)* (31 SNPs), superb starlings (*Lamprotornis superbus*) (Weinman *et al.,* 2014) (102 SNPs) and black-throated blue warblers (*Setophaga caerulescens*) (Kaiser *et al*., 2017) (97 SNPs). The first parent non-exclusion probabilities for the penguin panel was 0.2126; $8.4x10^{-5}$ for the starling panel and $1.9x10^{-2}$ for the warbler panel.

The development of a SNP-based parentage verification panel that could be applied across all the species of *Agapornis* would be preferable, in particular those species popular in aviculture. In a previous study, the *de novo* genome of *A. roseicollis* was sequenced, assembled and annotated (van der Zwan e*t al.*, 2018). This species is the only member of the intermediate group and displays characteristics of both other groups. Therefore, SNPs discovered in this species could possibly be representative of all the domesticated species. This was the first study to sequence the genome of any of the species in this genus. Currently, there is no SNP-based parentage verification test available for any of the *Agapornis*, or even parrot, species and development of such a test would aid parrot conservationists and breeders globally to verify pedigree and individual data.

The aim of this study was, therefore, to develop a SNP-based parentage verification panel for the most popular domesticated lovebird species. In this paper we describe the approach followed to discover SNPs by whole genome resequencing (WGR) of the reference genome individual's two parents. The variant caller Genome Analysis Toolkit (GATK) was used to identify and filter SNPs to be included in a parentage verification panel. SNPs were genotyped in a population of 960 lovebirds to determine the optimum number of SNPs needed to accurately infer parentage.

Materials and Methods

The genome of a young *A. roseicollis* male was previously sequenced, assembled and annotated (van der Zwan *et al.*, 2018) (NCBI accession number NDXB01000000) and used

here as the reference genome. After ethical approval was obtained from the North-West University AnimCare committee (Ethics number NWU00348-15-S5), 400 µL blood from each of the reference bird's parents were collected by a veterinarian in an EDTA tube. Sequencing was performed by Eurofins, Germany. Genomic DNA was isolated from the blood sample using the Machery Nagel Blood Mini kit according to the manufacturer's protocol with one modification, namely that 10 µl blood diluted with 190 µl PBS was used as starting material. Two shotgun libraries were constructed for each parent consisting of fragment sizes of 300 kb and 550 kb, respectively (van der Zwan *et al*., 2018). The libraries were sequenced on the Illumina HiSeq 2000 platform and sequencing was performed at a depth of 30x coverage for both birds (NCBI SRA accession number PRJNA355979).

The guidelines as set out by The Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010) was followed to discover variants from the sequencing data of both parents. All command line arguments can be viewed in Supplementary File 1. Raw variants including SNPs, Indels (insertions or deletions) and other variants were identified. SNPs located in indels should be excluded from a parentage verification panel (Heaton *et al.*, 2014) therefore indels were removed by applying hard filtering. The parameters and values applied during hard filtering are given in Table 1.

Table 1: Parameters and values applied in GATK during hard filtering of raw variants

Parameter	Value	Description
Quality score (Qual)		Phred-based probability of a false positive variant.
QualityByDepth (QD)	>2.0	Quality score adjusted for depth.
Fischer Strand (FS)	<10.0	Phred-scaled probability used to correct for sequencing bias using the Fisher's exact test.
RMSMappingQuality (MQ)	>50.0	The mapping quality of all the reads at that site.
MappingQualityRankSumTest (MQRankSum)	>-5.0	Comparison of the mapping qualities of the reads of the reference vs alternative allele.
ReadPosRankSumTest (ReadPosRankSum)	<-8.0	Indicates whether the position of the alternative and reference alleles is the same within the reads.

In order to identify SNPs that were found in both parents' genomes, a combined genotype file was created. The set of SNPs identified from the combined genotype file was sorted based on their QD scores and subsequently on their Quality (QUAL) scores. Most true heterozygote variants have a QD of approximately 12.0 (https://www.broadinstitute.org/gatk), therefore, only variants with QD scores between 11.5 and 12.5 were included in the dataset. Since areas with high sequencing coverage are likely to contain variants with high QUAL scores, QD compensates for this by normalizing the variant confidence by depth (https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_annotator_QualByDepth.php). The SNP with the highest QUAL score (with a QD score between 11.5 and 12.5) per scaffold was selected. This was done to ensure that SNPs were not located close to one another and possibly in linkage disequilibrium (Heaton *et* al., 2014). These variants were manually assessed in Integrated Genomics Viewer v 2.3.81 (IGV) (Robinson *et al*., 2011) by mapping the parent's reads to the reference genome. Each SNP was verified to ensure that it mapped to the reference genome, displayed accurate Mendelian inheritance and that they were only bi-allelic (Heaton *et al.,* 2014). Two custom array plates containing 240 SNPs each, were designed and manufactured by Thermo Fisher Scientific (Waltham, United States of America) to genotype 960 individuals on the QuantStudio 12K Flex platform.

To validate the NGS genotypes of the discovered SNPs, ten SNPs were randomly selected and Sanger sequencing was performed on DNA from the father and genome individual (Chick). Due to the nature of the assembly, only one allele per nucleotide was known for the Chick. No additional DNA could be obtained for the mother. Two additional individuals, one *A. fischeri* and one *A. taranta* were also sequenced to verify cross-species amplification. Sanger sequencing was performed by Inqaba Biotechnical Industries in South Africa. A comparison between the NGS and Sanger sequencing data is presented in Supplementary File 2.

Ethical clearance was obtained (NWU-00348-15-A5) to collect 960 biological samples from South Africa, Namibia, Belgium, The Netherlands and Spain from seven different lovebird species. The father and genome chick samples were also included to verify the genotypes. A total of 630 of the samples were collected in family structures consisting of one (either mother or father) or both parents and their chick(s) while the remaining 330 were randomly selected samples that were previously tested at a local laboratory (Lumegen laboratories) for sex determination. The number of families for each species and the different familial relationships are presented in Table 2.

Table 2: Number of samples per species tested and the number of one and two parent families included in the parentage verification analyses.

Species	Number of samples	One parent	Two parents
Agapornis canus	7		
Agapornis taranta	17		
Agapornis nigrigenis	32		3
Agapornis lilianae	34	1	
Agapornis personatus	72	1	2
Agapornis fischeri	298	7	12
Agapornis roseicollis	500	8	9
Total number of samples	960	17	26

DNA was extracted from dried blood or feather samples by Lumegen laboratories, South Africa following the manufacturer's protocol of the MagMax™ DNA multi-sample kit (Thermo Fisher Scientific) with only one modification namely that the blood cards were incubated overnight at 65˚C and the feathers at 55˚C. DNA samples were quantified using the Qubit Fluorometric quantification and NanoDrop methods (Thermo Fisher Scientific). DNA concentrations from especially the feather samples were low and a pre-amplification reaction for samples with concentrations below 15ng/µl was performed. Genotyping of the 960 samples at 480 SNPs was performed using the QuantStudio 12K Flex Real-time PCR system OpenArray technology (Thermo Fisher Scientific) at the Quantstudio 12K Flex Platform located at the University of Pretoria, South Africa. SNP data was analysed using the QuantStudio 12K Software as well as TaqMan® Genotyper software (both from Thermo Fisher Scientific). The father's genotype was used as the reference genotype since he was heterozygous at each SNP.

The father's sample did not amplify at 218 of the SNPs due to no amplification in one of the array plates where the sample was loaded. The remaining 262 SNPs where the father's genotype could be utilized as a reference, were assessed for inclusion. Observed Heterozygosity (H_O) (per SNP) (Morin *et al.*, 2004; Weinman *et al.*, 2014; Kaiser *et al.*, 2016), Minor Allele Frequency (MAF) (per SNP) (Heaton *et al.*, 2014; Talenti *et al.,* 2016), the number

of alleles, Hardy-Weinburg Equilibrium (HWE) (Liu *et al., 2016*) and the mean expected heterozygosity of the panel was used to evaluate the panel using Cervus 3.0 (Marshall *et al.*, 1998). The mean heterozygosity value of the 262-SNP panel was evaluated per species and in the total population.

Three different panels were constructed to determine the optimum number of SNPs to include in the panel, as well as the effect of MAF and H_O per SNP. MAF and H_O values are shown in Supplementary File 3. The first panel consisted of the total number of 262 SNPs. In the second panel all non-informative SNPs with H_O and MAF values below 0.1 were discarded, resulting in a panel of 195 SNPs. Finally, only SNPs with H_O and MAF exceeding 0.3 were included totalling 40 SNPs. Morin *et al.* (2004) as well as Heaton *et al.* (2014) recommended this threshold of 0.3 as the optimum value for inclusion of SNPs in a parentage verification panel. The robustness of the three panels were tested by verifying the pedigree data of 43 lovebird families, using Cervus 3.0 (Marshall *et al.,* 1998).

Results and Discussion

This was, to our knowledge, the first attempt to develop a parentage verification panel for *Agapornis* by discovering SNPs through WGR of the reference genome individual's parents. Sequencing of the parents yielded 44 923 Mbp of data for the father and 69 001 Mbp for the mother. In Table 3 the number of raw variants and SNPs (after hard filtering) identified during the GATK analyses for both parents and the combined file, is given.

Table 3: Variants and SNPs discovered for the mother, father and combined genotype file

	Mother	Father	Combined
Raw variants	2 156 950	1 601 584	N/A
SNPs only	1 916 289	1 428 869	1 667 639

Raw variants, including indels, were excluded during the variant calling phase of the study. For the mother, 240 661 and for the father 172 715 variants other than SNPs, were removed from the callset. The combined file was only created after hard filtering so no raw variants were called. A total of 1 667 629 SNPs were discovered in the combined genotype file including SNPs shared between the parents. This implies that the majority of the SNPs identified were shared between the parents, indicating that these SNPs might be found in the greater *A. roseicollis* population since the Mother and Father were unrelated. Hard filters were applied and a total of 103 287 SNPs passed the criteria. SNPs with QD scores between 11.5

and 12.5 amounted to 4 459, from where the subset of 480 SNPs, based on QUAL scores, were selected.

The genotypes of ten randomly selected SNPs that were included in the 480-SNP panel were validated by Sanger sequencing and the results are shown in Supplementary File 2. Since only one allele per nucleotide was known for the reference, both alleles couldn't be verified for the Chick. It was expected that the Father should be heterozygous at all SNPs since his genotype was verified during SNP identification. Only one discrepancy between the NGS and Sanger data was detected at SNP_951 where the Father was found to be a heterozygote in the NGS data but a homozygote in the Sanger data. It was later found that this SNP amplified poorly across all individuals during the QuantStudio 12K analyses and it was discarded from the study. Both the *A. taranta* and *A. fischeri* individuals amplified at all SNPs, except SNP_951. Both these individuals were homozygous across all nine amplified SNPs, but since the heterozygosity of these birds were unknown and the sub-set of SNPs very small, no final conclusion on the performance of the SNPs across species could be made. The comparison between the NGS and Sanger data showed that the NGS data was adequately sequenced at 30x coverage to discover SNPs correctly since the genotypes of the father was, for the most cases, identical at the ten SNPs analysed. It also showed that SNPs amplified across species, even though only homozygous alleles were detected.

In Table 4 the mean heterozygosity values of the total population (960 birds) as well as per species for the 262 panel is given. The Transitional and Sexually dimorphic groups (with the exception of *A. taranta*, 0.27) had mean heterozygosity values greater than 0.35 compared to the species in the White eye-ring group where all values were below 0.21.

Table 4: Mean expected heterozygosity at 262 SNPs typed

		Transitional group	Sexually dimorphic group		White eye-ring group			
Mean expected heterozygosity	All samples	*A. roseicollis*	*A. canus*	*A. taranta*	*A. personatus*	*A. lilianae*	*A. fischeri*	*A. nigrigenis*
	0.34	0.40	0.35	0.27	0.21	0.18	0.11	0.09

It has to be noted that only seven *A. canus* and seventeen *A. taranta* birds were included in this study and that more than half of the total number of samples were *A. roseicollis* samples, which could cause a bias in the heterozygosity results. Preferably, species or group-specific SNPs with high heterozygosity levels should be included in addition to the SNPs proposed in

this study to accommodate heterozygosity levels of all lovebird species. To date, however, only the *A. roseicollis* genome has been sequenced making SNP identification in other species challenging. Morin *et al.* (2004) and Kaiser *et al.* (2016) recommend a mean expected heterozygosity value of greater than 0.30, which was estimated in this study in the whole population (0.34).

The first, second and parent pair combined exclusion probabilities for each of the three panels were calculated and are given in Table 5. The 262 and 195-SNP panels had higher probabilities to exclude non-biological parents, compared to the 40-SNP panel (99.99% for both panels vs. 99.31% for the 40-SNP panel). The Second parent exclusion probability exceeded 99.99% for the 262 and 195 SNP panels but was slightly lower at 99.96% for the 40 SNP panel. Parent pair exclusion probabilities were similar (>99.9%) across all three panels. This compared well with panels such as the 95 SNPs of the trout panel (Liu *et al.*, 2016) (99.2%); 99.7% using 60 SNPs in pigs (Rohrer *et al.,* 2007) and 99.9% for a panel of 109 SNPs in sheep (Heaton *et al.*, 2014).

Table 5: The Mean Expected Heterozygosity values and Exclusion probabilities in the three different panels.

	SNP Panel size		
Exclusion probability	262	195	40
First parent	0,999999990	0,99999980	0,9931058
Second parent	0.9999999999	0.9999999999	0,99969307
Parent pair	0.9999999999	0.9999999999	0,99999753
Mean H_E	0.34	0.37	0.48

Three panels consisting of 262-SNPs, 195-SNPs and 40-SNPs were analysed in the parentage verification phase of this study. No species-specific panel was constructed. In Table 5 the mean Heterozygosity of each of the panels are shown. The exclusion of all non-informative SNPs with a H_O and MAF score below 0.1 had a small, but positive effect on the mean heterozygosity of the panel. As expected, the inclusion of only SNPs with scores in excess of 0.3 had a large effect on the mean H_E. The effect of the reduction in SNPs on parentage allocation will be discussed using the parentage verification analysis in the next section.

The 43 families with pedigree data as received from the breeders were analysed to verify their parentage in the 262-SNPs, 195-SNPs and 40-SNP panels. One family (*A. lilianae* family) amplified at too few loci for comparison between the chick and parent, and was discarded. Complete parentage verification results and relationships as received from the breeders are shown in Supplementary File 4. In Table 6 a summary of the parentage verification results with special focus on differences between the allocations of the three panels, is given.

Table 6: Summary of parentage allocations between the 262-SNPs, 195-SNPs and 40-SNP panels

	262 panel				195 panel				40 panel			
	No	‡Loci comp	#Mism	LOD scores	No	Loci comp	Mism	LOD scores	No	Loci comp	Mism	LOD scores
Families not allocated	10	125 - 261	12 – 60 (Pair) 24 – 83 (Trio)	All negative	10	94 – 194	11 – 31 (Pair) 17 – 61 (Trio)	All negative	6	8 – 38	1 (Pair) 1 – 7 (Trio)	All negative
§Alloc at *	30	102 - 260	0 – 6 (Pair) 2 – 19 (Trio)	All positive	30	62 – 194	0 - 5 (Pair) 0 – 12 (Trio)	All but one family positive	19	13 – 35	0 – 1 (Pair and trio)	All positive
flAlloc at +	2	30 and 71	2 - 7 (pair) 5 and 9 (Trio)	-1.40 and -9.67	2	44 and 51	1 (Pair) 4 and 5 (Trio)	Pair positive Trio negative	14 (Pair) 5 (Trio)	20 – 34	0 – 1 (Pair) 0 – 3 (Trio)	Negative and positive

‡ Loci comp: Number of loci compared between chick and parent.
# Mism: Genotypic mismatches between chick and one parent (pair) or two parents (trio).
§ Alloc at *: Parentage allocation made at the 95% strict confidence interval.
fl Alloc at +: Parentage allocation made at the 80% relaxed confidence interval.

Table 6 shows that as the number of SNPs were reduced, the parentage allocations changed. This is mainly due to a reduction in the number of SNPs compared between parents and offspring, leading to a reduction in the number of genotypic mismatches. Ten of the 43 (23.2%) families were not allocated as true parents when the 262-SNP and 195-SNP panels were analysed, highlighting the need for this study. This number decreased to 14% (6 families) when the 40-SNP panel were applied. The allocations of two families changed from the 80% relaxed confidence level during the 252-SNP panel analysis to the strict 95% level applying the 195-SNP panel. One family's allocation remained the same but the LOD score changed from positive (262-SNP panel) to negative (195-SNP panel).

Kaiser *et al.* (2016) (on the black throated blue warbler *Setophaga caerulescens*) stated that a panel of 40 SNPs with a mean H_O of 0.37 had the same power to assign parentage as a panel of 97 SNPs. Liu *et al*. (2016) (on rainbow trout *Oncorhynchus mykiss*) reported that a panel of 95 SNPs with a mean MAF of 0.34, allocated 100% parentage combinations correctly compared to poorer performance when using panels of 68 (99.2%), 48 (98%) and 36 (92.5%)

SNPs. Weinman *et al.* (2015) (on superb starlings (*Lamprotornis superbus*)) used 60 SNPs with a mean heterozygosity of 0.41 to accurately assign parentage and found only 1% incorrect allocations when using the 50 most heterozygous SNPs. In none of these studies, however, the effect of genotyping errors leading to genotypic mismatches were taken into account. The exclusion probability (P_E) of a panel of markers is calculated using the formula as described by Jamieson & Taylor (1997) and relies thereon that a conflict of alleles (or genotypic mismatch) at a specific locus between two individuals excludes the possibility of the supposed relationship between the two individuals (Baruch & Weller, 2008). The P_E value of a panel is directly influenced by the MAF of the SNPs as well as the genotyping error rate (Heaton *et al.*, 2014). If the error rate is high, there will be more mismatches and if the MAF is low genotypes will be more similar and exclusion not possible (Heaton *et al.*, 2014; Talenti *et al.*, 2016). In the current study (see Table 6 and Supplementary File 4) it was seen, especially in the 40-SNP panel, that by reducing the number of SNPs the number of mismatches automatically reduced. This resulted in more (false positive) allocations due to a higher LOD score. As these allocations could be compared to the other two panels this indicated false positive allocations and it was clear that the number of SNPs were too low to accurately exclude non-parents.

The number of SNPs required in a panel is strongly influenced by the mean expected heterozygosity of the panel and the individual SNP's H_O value (Morin *et al.,* 2004; Weinman *et al.*, 2015; Kaiser *et al.*, 2017). Liu *et al*. (2016) concluded that by using a larger number of SNPs, the allocations become less sensitive to MAF values and genotyping errors. This is further stressed by Strucken *et al.* (2016) who reports that the number of markers, rather than the quality of markers is a determining factor of success of a parentage test, and that at least 200 SNPs should be included when one parent's genotype is known. This is due to the fact that the exclusion probability formula doesn't take metrics such as MAF, H_O and genotypic error into account. Therefore, if the number of markers (and ultimately the number of mismatches) are reduced, the allocations could lead to false negative or false positive allocations. In this study the use of 40 highly informative SNPs (MAF and H_O >0.3) had less exclusion power and led to false positive and/or false negative allocations, compared to the full panel of 262 SNPs and the reduced 195-SNP panel with H_E and MAF values >0.1. This was confirmed in two families where only 70 (using the 262 SNP panel) and 30 (using the 195 SNP panel) SNPs were compared. The families were allocated at the 95% and 80% confidence intervals, respectively, despite having negative LOD scores, indicating false positive allocations.

The main selection criterion lovebird breeders use is that of Mendelian-inherited colour variations for which the genes and polymorphisms are yet to be discovered. This leads to

breeders often mating close relatives to ensure that recessive alleles are inherited. As the level of inbreeding might be very high in some populations (resulting in low H_O and MAF estimates), a sufficient number of SNPs should be included in the panel to accommodate this. The mean heterozygosity levels as well as the parentage allocations were similar in the 262 and 195-SNP panels. A reduced panel is more economical to genotype, and it is expected that the 195-SNP panel will be robust enough to exclude all non-parents in a population of various lovebird families and species. Genotypic errors could lead to false positive or negative allocations and high quality DNA samples should be used. Higher DNA concentrations should also result in higher amplification success, resulting in more loci to be compared. The SNPs were not highly heterozygous in the white eye-ring group species and therefore it is advised to identify SNPs unique to these species' genomes, in order to develop a panel that is highly heterozygous in all the domesticated lovebird species. Until such genomes become available, the SNP panel developed in the current study will be robust enough to exclude most non-parents in all *Agapornis* species.

CHAPTER FIVE: CONCLUSIONS AND FUTURE PROSPECTS

In this chapter conclusions of the main findings and results of this study (presented in Chapters 2-4) and their implications for lovebird breeding is presented. Future prospects for molecular screening tests that could be developed for this genus are also discussed.

5.1 SUMMARY AND CONCLUSION

There are nine extant species in the genus *Agapornis* found in their natural habitats across Africa and Madagascar and also bred and kept as pets across the globe. Breeders mainly select birds on the basis of plumage coloration, since over 30 plumage colour variations are found amongst these species, which are all inherited as Mendelian traits (as described in Chapter 2 and in the Published Paper I, Appendix A). Unfortunately, the genes of most of these plumage variations have not yet been identified and the identification thereof was beyond the scope of this study. The sequencing, assembly and annotation of the *A. roseicollis* genome in this study will however aid in the identification of these genes linked to the different colour variations. Until such, the parentage verification test developed during this study will aid breeders and buyers in ensuring that pedigrees are correct and that possible heterozygote offspring are selected as breeding stock. Since there was until this study no SNP-based genetic parentage verification test available specifically for parrot species, let alone *Agapornis* species, lovebird registration bodies might start to require genetic pedigree testing before a sale is made or a bird is imported or exported.

In order to address the issue of a lack of genetic screening tests, especially a parentage verification panel, the first reference genome of *Agapornis roseicollis* was sequenced, assembled and annotated (Chapter 3). This enabled the identification of SNPs that could be included in a parentage verification panel (Chapter 4).

5.1.1. SEQUENCING, ASSEMBLY AND ANNOTATION OF THE *AGAPORNIS ROSEICOLLIS* REFERENCE GENOME

One of the limitations to developing genetic screening tests for any species is the availability of a reference genome from where genes, markers and other genomic elements can be identified. Projects like Bird 10K aiming to sequence over 10 000 bird genomes are advancing well (Zhang *et al.,* 2015) and parrot genomes like the budgerigar, scarlet macaw, Puerto Rican parrot and the blue fronted Amazon has been sequenced and assembled, but thus far no African parrot (including lovebird) genomes have been sequenced. Amongst the over 350 parrot species, only twenty species are native to the African continent (Perrin, 2012), classified

as large parrots, small parrots and lovebirds. The first aim of this study was to sequence, assemble and annotate the genome of one of the *Agapornis* species, *A. roseicollis*. The results of this aim were published in Published Paper II (Draft *de novo* genome sequence of *Agapornis roseicollis* for application in avian breeding, Animal Biotechnology (2018) 29:4, 241-246). This species is phylogenetically classified as the "intermediate group" as it displays characteristics of both the sexually dimorphic group (*A. taranta, A. pullarius* and *A. canus*) as well as the white eye ring group (*A. fischeri, A. personatus, A. nigrigenis* and *A. lilianae*) (Dilger, 1960; Eberhard, 1998) (see Chapter 2 for more detail). This was the first study to sequence the genome of any of the African parrots or lovebird species, which gave a unique look at the structure of the genome of this sub-group of birds.

The genome of one young *A. roseicollis* male was sequenced at 100x coverage on the Illumina HiSeq 2000 platform using three shotgun sequence libraries (300, 550 and 750 bp, respectively) and two long jumping distance (LJD) paired-end libraries (3 Kbp and 8 Kbp). The assembly indicated the genome size was 1.1 Gbp and 15 045 coding gene sequences were annotated. The number of core genes present in the assembly was higher than the budgerigar, Puerto Rican parrot or scarlet macaw genomes. This confirmed that the genome was assembled accurately and to a completion level that will allow the development of genetic screening tests. The annotation of the genome also indicated that the number of genes identified compared well with other avian genomes, indicating that the annotation was correct and accurate.

The Scaffold N50 lengths were not as high as in other studies and can be improved by adding additional LJD libraries of insert sizes of 12 Kbp and/or 20 Kbp as performed by Zhang *et al* (2014). Another option to improve on this metric is to add additional data from another platform such as PacBio, which generates longer reads. However, given that the aim of this study was not genome completeness but rather to identify SNPs and, possibly in future studies, genes and other genomic elements for inclusion in genetic screening tests, the genome assembled up to this level was adequate.

5.1.2. WHOLE GENOME RESEQUENCING (WGR) TO DISCOVER SNPS AND GENOTYPING OF SNPS TO COMPILE A PARENTAGE VERIFICATION PANEL.

The second aim of this project was to discover SNPs throughout the genome after the reference genome was established. In this study one of the major contributions was to compile a SNP-based parentage verification panel for lovebirds. To achieve this, the parents of the reference genome individual were sequenced to identify SNPs by comparing their reads to the reference genome. The variant caller Genome Analysis Toolkit (GATK) was used to

identify SNPs throughout the genome. These SNPs were filtered based on parameters set out by GATK and a panel of 480 SNPs that passed these filters were included in the parentage verification panel.

WGR of the parents ensured that the SNPs followed a Mendelian inheritance pattern, and mapped to the reference genome, both of which were important in the development of a parentage verification panel (Heaton *et al.*, 2014; Talenti *et al.*, 2016; Liu *et al.*, 2016). By selecting one SNP per scaffold it was ensured that SNPs were not located too close to each other and weren't in linkage disequilibrium (Heaton *et al.*, 2014; Talenti *et al.*, 2016). Using the parents also allowed the identification of SNPs where both parents were heterozygotes. As only two birds were used to compile the initial panel of SNPs it was impossible to predict the level of heterozygosity of the SNPs in the wider lovebird population. By selecting SNPs that were heterozygous for both parents, the probability of increasing the level of heterozygosity of the total panel was greater. The use of GATK allowed the discovery of 1.6 million SNPs which proofed to be adequate in the development of a parentage verification panel.

The panel of 480 SNPs was genotyped in a larger population of 960 lovebirds in order to compile a SNP-based parentage verification panel. Genotypes of 262 SNPs were compiled as the first parentage verification panel. In order to determine the optimal number of SNPs to include in the panel, the number of SNPs were reduced to, firstly, include all SNPs with a H_O and MAF value of >0.1, resulting in 195 SNPs and, secondly, only SNPs with H_O and MAF values >0.3 were included resulting in a panel of 40 SNPs. Lastly, to determine the number of SNPs to include in the final panel, the three panels were used to verify parentage information of 43 lovebird families with known pedigree data, across the species as sent in by the breeders. The effect of incorrect pedigree data was seen when the proposed panel of SNPs identified that pedigree data as received from the breeders were incorrect for ten out of the 43 families that was used during the parentage analysis, spanning across species and aviaries.

All samples were sent in by breeders and consisted of feathers or dried blood stored on blood cards. Even though a DNA extraction kit was used that specifically allowed for feather DNA extraction, the DNA concentrations were very low. All samples with low concentrations were pre-amplified, but the result was still that many samples did not amplify well across the SNPs. The effect was that a low number of SNPs could be compared in some families which ultimately led to false parentage allocations. Poor quality DNA is also one of the causes of genotyping errors which, in turn, caused genotyping mismatches and false allocations during the parentage verification phase. The number of loci compared were also lower in some cases which led to incorrect parentage allocation results (see Submitted Manuscript I for more detail).

The panel of 262 SNPs were evaluated based on the mean expected heterozygosity (H_O) value of the whole population and per species. Since the SNPs were not as polymorphic in the white-eye ring group compared to the other species, it would be advisable to, once one of the genomes of these species are available, select SNPs identified specifically for these species. These SNPs could be added to the current proposed panel or offered as an additional panel. The proposed panel did, however, allocate all parentage cases of all species tested accurately and therefore this panel should be adequate to verify parentage until such genomes are available.

In summary, the variant caller and genotyping platform were both excellent choices for SNP discovery and genotyping. The poor DNA concentration had a negative effect on the number of SNPs that were genotyped and led to those samples having a very low number of genotypes available to compare between parents and chicks. The panel of 195 SNPs was robust enough with a high mean heterozygosity level and exclusion probability to accurately exclude non-parents. It was however found that in two families where less than 70 SNPs were typed and compared between chick and parent, the parentage allocation was incorrectly assigned. This was due to the fact that less SNPs were compared and less mismatches were detected. Most of these mismatches were most likely caused by the poor DNA concentration due to genotyping errors. The panel of 195 SNPs is proposed as a parentage verification panel for lovebirds.

5.2 CONCLUSION ON CONTRIBUTIONS MADE BY THIS STUDY

In summary, this was the first study to sequence, assemble and annotate the genome of any of the lovebird species, which was the first aim of this project. This reference genome will act as the blueprint from where SNPs and genes of economic importance can be identified in future studies. The sequencing at 100x coverage using the Illumina HiSeq 2000 platform proved to yield sufficient data to assemble the genome up to scaffold level. The addition of 12 Kbp and 20 Kbp LJD libraries or the addition of PacBio reads might aid in creating longer Scaffold N50 lengths for future projects. However, for this project where the main use of the genome was to identify SNPs to include in a parentage verification panel, this assembly was sufficient. The second and third major aims of this study was to identify SNPs throughout the genome in order to develop a parentage verification panel for domesticated lovebird species. The use of the parents to identify SNPs proofed to be valuable as the inheritance of the SNPs could be confirmed and validated against the reference genome. Over 1 600 000 SNPs were identified throughout the combined genomes of the mother and father using GATK. GATK was powerful enough to identify high quality SNPs that were amplified across all domesticated

lovebird species. A population of 960 lovebirds were genotyped at 480 selected SNPs to determine the optimal number of SNPs to be included in the lovebird parentage verification panel. After genotyping, a panel of 262 SNPs was constructed that was further reduced based on the mean heterozygosity of the panel, expected heterozygosity (H_O) and minor allele frequency (MAF) per SNP in order to determine the least number of SNPs with the highest exclusion power. It was determined that a panel of 195 SNPs with a mean heterozygosity of 0.378, H_O and MAF values (per SNP) >0.1 and an average first parent exclusion probability of 99.9% was adequate to accurately determine parentage in 43 lovebird families across most domesticated lovebird species. This panel was proposed as a commercial parentage verification panel for all lovebird species. This panel can be applied in verifying pedigree data and individual identification in breeding systems and also in conservation efforts. It will also help reduce the number of fraudulent transactions where heterozygous offspring with wildtype plumage colour are sold without any proof of pedigree or colour genotype. This study clearly shows the benefits of *de novo* sequencing of a non-model organism genome, especially in species used for production or companionship as SNPs and genes of economic importance can be discovered from the reference genome.

5.3 FUTURE PROSPECTS

This was the first study where a molecular genetic screening test for any of the *Agapornis* species was developed. To date, most of the genome assembly studies on parrots were focused on conservation efforts and not the application in aviculture. The development of a SNP-chip for all traits of economic importance was not within the scope of this study, but the sequencing, assembly and annotation of a reference genome would act as the blueprint from which all future studies could follow. If a single SNP-chip based test could be developed for lovebird breeding, it will enable breeders to test a bird once and based on the results, select birds based on their genotype on a wide range of traits. These traits include parentage (as developed in this study), plumage colour genotype, gender, species and even health markers like immune response. Due to the fact that the reference genome of one of the species are now available, the development of these tests are feasible. The prospects to develop a SNP-chip with traits of economic importance for all *Agapornis* species are discussed below.

IMPROVEMENT OF THE PARENTAGE VERIFICATION PANEL ACROSS WHITE-EYE RING GROUP SPECIES

In Chapter 4 (Submitted Manuscript I) it was shown that the species from the white-eye ring group, in particular *A. fischeri*, *A. nigrigenis* and *A. lilianae*, performed poorly during the mean heterozygosity analysis (see Table 4 in Chapter 4). Since *A. fischeri* and *A. personatus* are

currently the two most popular species to breed with, the assembly and subsequent SNP identification from these genomes would aid in improving the lovebird parentage verification panel. The current proposed parentage verification panel was however capable of excluding non-parents across six species (including *A. fischeri* and *A. nigrigenis*).

The SNP-based parentage test developed for *Agapornis* species in this study could be expanded to other parrot species frequently found in aviculture, such as budgerigars and ringnecks (*Psittacula krameri*), to verify if these SNPs could be applied for parentage verification across the Psittacidae family. Since it is unknown if these SNPs are found in other species, a range of individuals from other genera will have to be tested in order to verify their polymorphism.

COLOUR GENOTYPING

The main economic trait breeders use to select birds is plumage colour variations and lovebirds display at least 30 different colour variations, some of which can also be combined. Parrots have a unique mechanism and pigment called psittacofulvins (Hill & McGraw, 2006) and very little molecular genetic research has been conducted on the genes and polymorphisms causing different colours. Thus far only one polymorphism resulting in blue, instead of the wildtype green coloration in budgerigars has been identified (Cooke *et al*., 2017) but hasn't been confirmed to cause the same phenotype in lovebirds. Through breeder's records it was established that most of these traits are inherited as simple Mendelian traits. These mechanisms and variations are described in more detail in Chapter 2 and in the Published Paper I, Annexure A. If the genes and polymorphisms linked to the variations are discovered and verified, these SNPs could be included in the SNP chip and breeders could test for all the colour genotypes in one single test. The possibility also exists to expand these tests to other parrot species such as the budgerigar and ringnecks as they too display many of the same colour variations.

SEXING

Due to the fact that the genders of five of the *Agapornis* species are monomorphic in colour, breeders have to confirm the bird's gender before it is placed with a mate during breeding season. Birds have a Z/W sex-chromosome system (as opposed to the X/Y system in mammals) and females are ZW whereas males are ZZ. The current test that is available tests for the CHD gene (Chromo Helicase DNA-binding gene) as it has different lengths on the Z and W chromosomes and therefore males and females can easily be distinguished with a molecular PCR-based test (Griffiths et al., 1996; Jensen et al., 2003). SNPs within or close to this gene or on other areas of the Z or W chromosomes could also be linked to gender and

could be included on the chip. Currently, no studies to identify SNPs linked to sex in parrots exist. By identifying these SNPs, breeders would not have to perform a separate PCR-based test to determine the gender of a bird.

SPECIES IDENTIFICATION

Many of the *Agapornis* species can intermate and mating between lovebirds and budgerigars producing viable young birds have also been recorded (Chapter 2 and Published Paper II, Annexure A). Since their habitats do not overlap (with the exception of *A. swindernianus* and *A. pullarius*) these hybridisations do not occur in the natural population. In aviculture, reports of genetic introgression to introduce a specific colour variation to another species is well known (e.g. the dark factor phenotype that was introduced from *A. roseicollis* to *A. personatus* and from there to other species as explained in Chapter 2 and the Published Paper I, Annexure A). Hybrids are not generally accepted at shows and auctions, but, unfortunately, genetic introgression is used as a tool to introduce new colours or even to breed new species. In addition, many inexperienced or ignorant breeders place different species in an open colony breeding system where birds are allowed to breed freely. These young hybrids are then sold on the pet market to unknowing owners. By identifying SNPs linked to specific species a speciation identification test could be developed. These SNPs could be added to the SNP-chip and allow prospective buyers to screen that a bird is pure for that species. This test can also be utilized by conservationists and custom officials to identify a species from a tissue or egg sample. This could reduce poaching of birds and expedite prosecution of offenders. This would however require the sequencing of other lovebird species genomes.

IMMUNITY

One of the biggest health challenges that parrots, and lovebirds are no exception, face, is Psittacine beak and feather disease (PBFD) caused by the beak and feather disease virus (BFDV). It is estimated that up to 10-20% of South African breeding parrot stock is lost due to this virus and it is a worldwide problem (Heath *et al*., 2004). Peters *et al.* (2014) describe the threat this virus has to natural populations, as in the case where the last wild population of the orange bellied parrot (*Neophema chrysogaster*) was infected by this virus and could be extinct. By investigating if some species, lines or families have a natural immunity against the virus, birds can be selected for their immunity and the disease can be limited. By identifying Quantitative Trait Loci (QTLs) through a Genome Wide Association Study (GWAS) genes and ultimately SNPs that are linked to natural immunity could be identified. A GWAS approach was followed by Luo *et al*. (2014) where SNPs linked to natural immunity against the bronchitis virus in chickens was identified. These SNPs can once again be included on the SNP chip

allowing breeders to select stronger birds. This study could also be expanded to other parrot species plagued by this virus.

REFERENCES

Baker, M. 2012. *De novo* genome assembly: what every biologist should know. *Nature methods*. 9:4, 333 - 337.

Barnes, T. 2017. EU wild bird import ban cuts trade worldwide by 90%. Independent. https://www.independent.co.uk/environment/eu-wild-bird-import-ban-trade-worldwide-cut-90-per-cent-animal-rights-species-populations-a8071861.html

Baruch, E. & Weller, J.I. 2008. Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics.* 39:5, 474 - 479.

Berg, M.L. & Bennett, A.T.D. 2010. The evolution of plumage colouration in parrots: a review. *Emu*, 110, 10 – 20.

Beuzen, N.D., Stear, M.J. & Chang, K. C. 2000. Molecular markers and their use in animal breeding. *The Veterinary Journal*. 160, 42 – 52.

Boisvert, S., Laviolette, F. & Corbeil, J. 2010. Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of computational biology*. 17, 1519 - 1533.

Bradnam, K.R. Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W-C., Corbeil, J., Del Fabbro, C., Docking, T.R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N.A., Ganapathy, G., Gibbs, R.A., Gnerre, S., Godzaridis, E., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J.B., Ho, I.Y., Howard, J., Hunt, M., Jackman, S.D., Jaffe, D.B., Jarvis, E.D., Jiang, H., Kazakov, S., Kersey, P.J., Kitzman, J.O., Knight, J.R., Koren, S., Lam, T-W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., MacCallum, I., MacManes, M.D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T.D., Paten, B., Paulo, O.S., Phillippy, A.M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F.J., Richards, S., Rokhsar, D.S., Ruby, J.G., Scalabrin, S., Schatz, M.C., Schwartz, D.C., Sergushichev, A., Sharpe, T., Shaw, T.I., Shendure, J., Shi, Y., Simpson, J.T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B.M., Wang, J., Worley, K.C., Yin, S., Yiu, S-M., Yuan, J., Zhang, G., Zhang, H., Zhou, S. & Korf, I.F. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, 2, 10.

Brush, A.H. 1990. Metabolism of carotenoid pigment in birds. *FASEB J*, 4, 2969 – 2977.

Buckley, PA. 1987. Avian genetics: a population and ecological approach. London: Academic Press.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C. & Jaffe, D.B. 2008. ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Research.* 18: 810 – 820.

Cai, Q., Qian, X. & Lang, Y. 2013. Genome sequence of ground tit *Pseudopodoces humilis* and its adaptation to high altitude. *Genome Biology*. 14: R29.

Chapman, J.A., Ho, I, Sunkara, S., Luo, S., Schroth, G.P. & Rokhsar, D.S. 2011. Meraculous: De novo Genome assembly with short pared-end reads. *PLOS One*. 6(8): e23501.

Choi, J-W., Chung, W-H., Lee, K-T., Choi, J-W., Jung, K-S., Cho, Y., Kim, N. & Kim, T-H. 2013. Whole genome resequencing of Heugu (Korean black cattle) for the genome-wide SNP discovery. *Korean Journal of Food Science of Animal Resources*. 33:6, 715 - 722.

Coetzer, W.G., Downs, C.T., Perrin, M.R. & Willows-Munro, S. 2017. Testing of microsatellite multiplexes for individual identification of Cape Parrots (Poicephalus robustus): paternity testing and monitoring trade. *PeerJ* 5:e2900.

Cooke, T.F., Fischer, C.R., Wu, P., Khosla, C., Chuong, C-M. & Bustamante, C.D. 2017. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell*. 171, 427 – 439.

D'Alba, L., Kieffer, L. & Shawkey, M.D. 2012. Relative contributions of pigments and biophotonic nanostructures to natural color production: a case study in budgerigar (*Melopsittacus undulatus*) feathers. *Journal of Experimental Biology* 215: 1272 – 1277.

Dalloul, R.A., Long, J.A., Zimin, A.V., Aslam, L., Beal, K., Blomberg, LA., Boufard, P., Burt, D.W., Crasta, O., Crooijmans, R.P.M.A., Cooper, K., Coulombe, R.A., De, S., Delany, M.E., Dodgson, J.B., Dong, J.., Evans, C., Frederickson, K.M., Flicek, P., Florea, L., Folkerts, O., Groenen, M.A.M., Harkins, T.T., Herrero, J., Hoffmann S., Megens, H-J., Jiang, A., de Jong, P., Kaiser, P., Kim, H., Kim, K-W., Kim, S., Langenberger, D., Lee, M-K., Lee, T., Mane, S., Marcais, G., Marz, M., McElroy, A.P., Modise, T., Nefedov, M., Notredame, C., Paton, I.R., Payne, W.S., Pertea, G., Prickett, D., Puiu, D., Qioa, D., Raineri, E., Ruffier, M., Salzberg, S.L., Schatz, M.C., Scheuring, C., Schmidt, C.J., Schroeder, S., Searle, S.M.J., Smith, E.J., Smith, J., Sonstegard, T.S., Stadler, P.F., Tafer, H., Tu, Z., Van Tassell, C.P., Vilella, A.J., Willims, K.P.,Yorke, J.A., Zhang, L., Zhang, H-B., Zhang, X., Zhang, Y. & Reed, K.M. 2010. Multi-Platform next-generation sequencing of the domestic Turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLOS Biology*. 8:9, e1000475.

Damas, J., O'Connor, R., Farré, M., Panagiotis, V., Lenis, E., Martell, H.J., Mandawala, A., Fowler, K., Joseph, S., Swain, M.T., Griffin, D.K. & Larkin, D.M. 2017. *Genome Research*. 27: 875 - 884.

Delacour. J. 1942. Bird breeding in France. *Avicultural Magazine*, 5th Series 7: 28.

DeNise. S., Johnston. E., Halverson. J., Marshall. K., Rosenfeld. D., McKenna. S., Sharp. T. & Edwards. J. 2003. Power of exclusion for parentage verification and probability of match for identity in American kennel club breeds using 17 canine microsatellite markers. *Animal Genetics.* 35: 14 – 17.

Denisov. G., Walenz. B., Halpern. A.L., Miller. J., Axelrod. N., Levy. S. & Sutton. G. 2008. Consensus generation and variant detection by Celera Assembler, *Bioinformatics*. 24:8, 1035 – 1040.

Diega. A. 2017. Bird flu to increase foreign demand for African lovebirds. Business Mirror, 21 May. Available at https://businessmirror.com.ph/bird-flu-to-increase-foreign-demandfor-african-lovebirds/ [accessed 18 April 2018].

De Summa, S., Malerba, G., Pinto, R. Mori, A., Mijatovic, V. & Tommasi S. 2017. GATK hard filtering: tuneable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinformatics.* 18 (Supplement 5):119.

Dilger, W. 1960. The comparative ethology of the African parrot genus Agapornis. *International Journal of Behavioral Biology Ethology*. 17:6, 649 - 685.

Dyck, J. 1971 (a). Structure and colour-production of the blue barbs of *Agapornis roseicollis* and *Cotinga maynana*. Zeitschrift fur Zellforschung und mikroskopische Anatomie. 115, 17 – 29.

Dyck, J. 1971 (b). Structure and spectral reflectance of green and blue feathers of the rose-faced lovebird (*Agapornis roseicollis*). *Det. Kongelige Danske Videnskabernes Selskab, Bilogiske Skrifter.* 18, 2.

Earl, D., Bradnam, K., St. John, J., Darlin, A., Lin, D., Fass, J., Yu, H.O.K., Buffalo, V., Zerbino, D.R., Diekhans, M., Nguyen, N., Ariyarantne, P.N., Sung, W-K., Ning, Z., Haimel, M., Simpson, J.T., Fonseca, N.A., Birol, I., Docking, T.R., Ho, I.Y., Rokhsar, D.S., Chikhi, R., Lavenier, D., Chapuis, G., Naquin, D., Maillet, N., Schatz, M.C., Kelley, D.R., Phillippy, A.M., Koren, S., Yang, S-P., Wu, W., Chou, W-C., Srivastava, A., Shaw, T.I., Ruby, J.G., Skewes-Cox, P., Betegon, M., Dimon, M.T., Solovyev, V., Seledtsov, I., Kosarev, P., Vorobyev, D., Ramirez-Gonzalez, R., Leggett, R., MacLean, D., Xia, F., Luo, R., Li, Z., Xie, Y., Liu, B., Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Yin, S., Sharpe, T., Hall, G., Kersey, P.J., Durbin, R., Jackman, S.D., Chapman, J.A., Huang, X., DeRisi, J.LL., Caccamo, M., Li Y., Jaffe, D.B., Green, R.E., Haussler, D., Korf, I. & Paten, B. 2011. Assemblathon 1: A competitive assessment of *de novo* short read assembly methods. *Genome Research*. 21: 2224 - 2241.

Eberhard, J.R. 1998. Evolution of Nest-Building Behavior in Agapornis Parrots. *The Auk.* 115:2, 455 - 464.

Egwumah, F.A. & Iboyi, M.O. 2017. Paramount factors influencing the breeding performance of lovebird *Agapornis pullaria*. *IJLST*. 10:7, 69 - 74.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution.* 29: 1, 51 - 63.

Elliot, F.S. 2013. Budgerigar matings and colour expectations. United Kingdom: Read Books.

Farner, DS, King, JR & Parkes, KC (eds). 1982. Avian biology, vol. 6. New York, United States of America: Academic Press.

Fels, L. & Distl, O. 2014. Identification and validation of quantitative trait loci (QTL) for canine hip dysplasia (CHD) in German Shepherd dogs. *Pols One*. 9:5, e96618

Fleming, D.S., Koltes, J.E., Fritz-Waters, E.R., Reecy, J.M. & Lamont, S.J. 2015. A tale of two birds: SNP discovery and genomic architecture of highly inbred Leghorn and Fayoumi chicken breeds using Whole Genome Resequencing. Proceedings, 10th World Congress of Genetics Applied to Livestock Production.

Flicek, P., Ahmed, I., Amode, R. & Barrell, B. 2012. Ensembl 2013. *Nucleic Acids Research*. 1:8.

Forshaw, J. 1989. Parrots of the World. United States of America: Princeton University Press.

Ganapathy, G. Howard, J.T., Ward, J.M., Li, J., Li, B., Li, Y., Xiong, Y., Zhang, Y., Zhou, S., Schwartz, D.C., Schatz, M., Aboukhalil, R., Fedrigo, O., Bukovnik, L., Wang, T., Wray, G., Rasolonjatovo, I., Winer, R., Knight, J.R., Koren, S., Warren, W.C., Zhang, G., Phillippy, A.M. & Jarvis, E.D. 2014. High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience*, 3, 11.

Garvin, M. R., Saitoh, K. & Gharrett, A.J. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources.* 10, 915 - 934.

Genome 10K Community of Scientists. 2009. Genome10K: A proposal to obtain whole-genome sequences for 10 000 vertebra species. *Journal of Hereditary*. 100:6, 659 - 674.

Gheyas, A.A., Boschiero, C., Eory, L., Ralph, H., Kuo, R., Woolliams, J.A. & Burt, D.W. 2015. Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research*. 22: 3, 205 - 217.

Griffiths, R., Daan, S. & Dijkstra, C. 1996. Sex identification in birds using two CHD genes. *Proceedings of the Royal Society Of London.* Series B: Biological Sciences.

Groenen, M.A.M., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L. B., Dodgeson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H. & Vignal, A. 2000. A Consensus linkage map of the chicken genome. *Genome Research*. 10: 137 - 147.

Guay, P-J., Potvin, D.A. & Robinson, R.W. 2012. Aberrations in plumage coloration in birds. *Australian Field Ornithology.* 29: 23–30.

Gunnarsson, U., Hellström, A.R., Tixier-Boichard, M., Minvielle, F., Bed'hom, B., Ito, S., Jensen, P., Rattink, A., Vereijken, A. & Andersson, L. 2007. Mutations in SLC45A2 cause plumage color variation in chicken and Japanese quail. *Genetics.* 17: 857 – 577.

Hampe, H. 1957. Die Unzertrennlichen. Pfungstadt, Germany.

Harris, R.L. 1979. Genetics and the budgerigar. *AFA Watchbird Journal of the American Federation of Aviculture.* 6: 2.

Hayward, J. 1979. Lovebirds and their Colour Mutations. United Kingdom: Blandford Press Ltd.

Heath, L., Martin, D.P., Warburton, L., Perrin, M., Horsfield, W., Kingsley, C., Rybicki, E.P. & Williamson, A-L. 2004. Evidence of unique genotypes of Beak and Feather Disease Virus in Southern Africa. *Journal of Virology*. 78:17, 9277 - 9284.

Heaton, M.P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., McEwan, J., Maddox, J.F., Basnayake, V., Petrik, D.T., Simpson, B., Smith, T. P. L., Chitko-McKown, C.G., & The International Sheep Genomics Consortium. 2014. SNPs for Parentage Testing and traceability in globally diverse breeds of sheep. *Plos One*. 9:4, e94851.

Hedrick, P.W. 2000. Genetics of Populations. 2nd Edition. Jones and Bartlett Publishers, Inc. United States of America.

Hill, G.E. & McGraw, K.J. (eds). 2006. Bird coloration, vol. 1: mechanisms and measurements. Cambridge, MA, United States of America: Harvard University Press.

Hillier, L.W. & International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 432: 695 – 716.

Hohenlohe, P.A., Amish, S.J. Catchen, J.M., Allendorf, F.W. & Luikart, G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular ecology resources*. 11, (Suppl. 1), 117 – 122.

Hubbard, J.K., Uy, J.A.C., Hauber, M.E., Hoekstra, H.E. & Safran, R.J. 2010. *Trends in Genetics*. 26:5, 231 - 239.

Hume, J.P. & van Grouw, H. 2014. Colour aberrations in extinct and endangered birds. *Bulletin of the British Ornithologists' Club.* 134: 168 – 193.

Israel, C & Weller, JI. 2000. Effect of misidentification on genetic gain and estimation of

breeding value in dairy cattle populations. *Journal of Dairy Science.* 83: 181 – 187.

Jamieson, A. & Taylor, St.C.S. 1997. Comparisons of three probability formulae for parentage exclusion. *Animal Genetics*. 28:6, 397 - 400.

Jan, C & Fumagalli, L. 2016. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). PeerJ 4: e2416.

Jensen, T., Pernasetti, F.M. & Durrant, B. 2003. Conditions for rapid sex determination in 47 avian species by PCR of genomic DNA from blood, shell-membrane blood vessels, and feathers. *Zoo Biology*. 22:6.,561 - 571.

Kaiser, S.A., Taylor, S.A., Chen, N., Sillett, T.S., Bondra, E.R. & Webster, M.S. 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources.* 17, 183 - 193.

Kardos, M., Husby, A., McFarlane, S.E., Qvarnström, A. & Ellegren, H. 2016. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*. 16:3, 727 - 741.

Koren, S., Schatz, M., Walenz, B., Martin, J., Howard, J., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, R., Jarvis, E.D. & Phillippy, A.M. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*. 30: 693 – 700.

Koutsogiannouli, E.A., Moutou, K.A., Sarafidou, T., Stamatis, C. & Mamuris, Z. 2010. Detection of hybrids between wild boars (*Sus scrofa scrofa*) and domestic pigs (*Sus scrofa f. domestica*) in Greece, using the PCR-RFLP method on melanocortin-1 receptor (MC1R) mutations. *Mammalian Biology.* 75: 69 – 73.

Kumar, S., Banks, T.W. & Cloutier, S. 2012. SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics*. Vol. 2012, Article ID 831460.

Labuschagne, C., Nupen, L., Kotzé, A, Grobler, P.J. & Dalton, D.L. 2015. Assessment of microsatellite and SNP markers for parentage assignment in ex situ African Penguin (*Spheniscus demersus*) populations. *Ecology and Evolution.* 5, 4389 – 4399.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 25:17, 54 - 60.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S & The Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:16, 2078 - 2079.

Liu, S., Palti, Y., Gau, G. & Rexroad, C. E. III. 2016. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture*. 452, 178 - 182.

Lou, C., Qu H., Ma, J., Wang, J., Hu, X., Li, N. & Shu, D. 2014. A genome-wide association study identifies major loci affecting the immune response against infectious bronchitis virus in chicken. *Infection, Genetics and Evolution*. 21, 351 - 358.

Luo, R, Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu,

S-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T-W. & Wang, J. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 1, 18.

Manegold, A. 2013. Two new parrot species (Psittaciformes) from the early Pliocene of Langebaanweg, South Africa, and their palaeoecological implications. *Ibis*. 155, 127 - 139.

Manegold, A. & Podsiadlowski, L. 2014. On the systematic position of the Black-collared Lovebird *Agapornis swindernianus* (*Agapornithinae*, Psittaciformes). *Journal of Ornithology*. 155, 581 – 589.

Marez, C. 2003. The Abyssinian lovebird Agapornis taranta. *AFA Watchbird Journal of the American Federation of Aviculture*. 30, 3.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology.* 7: 639 - 655.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connoll, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S. & van Tassell, C.P. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLOS One*. 4: 4, e5350.

McCarthy, E. 2006. Handbook of Avian Hybrids of the World. Oxford University Press, United Kingdom.

McGraw, K.J. & Nogare, M.C. 2004. Distribution of unique red feather pigments in parrots. *Biology Letters.* 1, 38 – 43.

McGraw, K.J. 2006. Mechanics of uncommon colors: Pterins, Porphyrins, and Psittacofulvins. In: Bird Coloration, Vol 1, Mechanisms and Measurements by Hill, G.E. & McGraw, K.J. (Eds.) Cambridge, MA, United States of America: Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabirel, S., Daly, M. & DePristo, M.A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20: 1297 - 1303.

Miller, J.R., Delcher, A.L., Koren, S., Venter, E., Walenz, B.P., Brownley, A., Johnson, J., Li, K., Mobarry, C. & Sutton, G. 2008. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*. 15:24, 2818 - 2824.

Moreau, R. 1948. Aspects of evolution in the parrot genus *Agapornis*. *IBIS*. 90: 2, 206 - 239.

Morin, P.A., Luikart, G., Wayne, R.K. & The SNP working group. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*. 19: 4, 208 - 216.

Morinha, F., Cabral, J.A. & Bastos, E. 2012. Molecular sexing of birds: a comparative review of polymerase chain reaction (PCR)-based methods. *Theriogenology*. 78:4, 703 - 714.

Mundy, N.I. 2006 (a). A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of the Royal Society B*. 272, 1633 – 1640.

Mundy, N.I. Genetic Basis of Color Variations in Wild birds. 2006 (b). In: Bird Coloration, Vol 1, Mechanisms and Measurements by Hill, G.E. & McGraw, K.J. (Eds.) Cambridge, MA, United States of America: Harvard University Press.

Mundy, N.I. 2018. Colouration genetics: Pretty polymorphic parrots. *Current biology.* 26: 3, R113 - R114.

Mwangomo, E.A., Hardesty, L.H., Sinclair, A.R.E., Mduma, S.A.R. & Metzger, K.L. 2007. Habitat selection, diet and interspecific associations of the rufous-tailed weaver and Fischer's lovebird. *African Journal of Ecology*. 46, 267 – 275.

Mzumara, T.I., Perrin, M.R. & Downs, C.T. 2016 (a). The drinking habits of the Lilian's lovebird and incidents of poisoning at waterholes. *African Journal of Ecology*. 54: 3, 289 – 298.

Mzumara, T.I., Perrin, M.R. & Downs, C.T. 2016 (b). Comparative performance of four survey methods for assessing Lilian's Lovebird abundance in Liwonde National Park, Malawi. *Ostrich*. 87: 3, 247 – 253.

Nanda, I., Karl, E.K, Griffin, D.K., Schartl, M. & Schmid, M. 2007. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenetic Genome Research.* 117, 43 – 53.

Ndithia, H., Perrin, M.R. & Waltret ,M. 2007. Breeding biology and nest site characteristics of the Rosy-faced Lovebird *Agapornis roseicollis* in Namibia. *Ostrich*. 78: 1, 13 – 20.

Neunzig, R. 1926. Zur Systematik und Biologie der Gattung *Agapornis*. Verhandlungen der Orni- thologischen Gesellschaft in Bayern. 17:112.

Oleksyk, T.K., Pombert, J-F., Siu, D., Mazo-Vargas, A., Ramos, B., Guiblet, W., Afanador, Y., Ruiz-Rodriguez, C.T., Nickerson, M.L., Logue, D.M., Dean, M., Figueroa, L., Valentin, R. & Martinez-Cruzado, J-C. 2012. A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. *GigaScience*. 1, 14.

Olsen, M.T., Volny, V. H., Bérubé, M., Dietz, R., Lydersen, C., Kovacs, K. M., Dodd, R.S. & Palsbøll, P.J. 2011. A simple route to single-nucleotide polymorphism in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (*Pusa hispida hispida*). *Molecular Ecology Resources*. 11 (Suppl. 1) 9 - 19.

Paszkiewicz, K. & Studholme, D.E. 2010. *De novo* assembly of short sequenced reads. *Briefings in Bioinformatics*. 11:5, 457 - 572.

Peters, A., Patterson, E.I., Baker, B.G.B., Moldsworth, M., Srker, S., Ghorashi, S.A. & Raidal, R. 2014. Evidence of Psittacine beak and feather disease virus spillover into wild critically endangered orange-bellied parrots (*Neophema chrysogaster*). *Journal of Wildlife Diseases*. 50:2, 288 - 296.

Perrin, M.R. 2009. Niche separation in African parrots. Pp. 29–37. In: Harebottle, D.M., Craig, A.J.F.K., Anderson, M.D., Rakotomanana, H. & Muchai, M. (eds). Proceedings of the 12th Pan-African Ornithological Congress, 2008. Cape Town, South Africa.

Perrin, M.R. 2012. Parrots of Africa, Madagascar and the Mascarene islands: biology, ecology and conservation. Johannesburg, South Africa: Wits University Press.

Pires, S.F. 2012. The illegal parrot trade: a literature review. *Global Crime*. 13:3, 176 - 190.

Prum, R.O., Morrison, R.L. & Ten Eyck, G.R. 1994. Structural color production by constructive reflection from ordered collagen arrays in a bird (*Philepitta castanea*: Eurylaimidae). *Journal of Morphology*. 222 (1), 61 - 72.

Prum, R.O. 2006. Anatomy, physics, and evolution of structural colours. In: Bird Coloration Mechanisms and measurements, Volume I. United States of America: Harvard University Press.

Qiu, J., Simpson, B., Kock L., Donner, J., Cole, C., Davison, S., Dunn, M., Bannasch, D. & Boyko, A. 2016. Evaluation of single nucleotide polymorphism (SNP) markers for canine parentage analysis. *Journal of Animal Science*. 94, 4:1, 106 – 107.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific iosciences and Illumina MiSeq sequencers. *BMC Genomics*. 13 : 341.

Rasmussen, R.S. & Morrissey, M.T. 2008. DNA-based methods for the identification of commercial fish and seafood species. *Comprehensive Reviews in Food Science and Food Safety.* 7: 280 – 295.

Rheindt, F.E. & Edwards, S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *The Auk.* 128: 620 – 632.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. 2011. Integrative Genomics Viewer. *Nature Biotechnology.* 29, 24 - 26.

Rohrer, G. A., Freking, B.A. & Nonneman, D. 2007. Single nucleotide polymorphisms for pig identification and parentage exclusion. *Animal Genetics*. 38, 253 - 258.

Sage, B.L. 1962. Albinism and melanism in birds. *British Birds.* 55: 6.

Schaefer, R.J., Schubert, M., Bailey, E., Bannasch, D.L., Barrey, E., Bar-Gal, G.K., Brem, G., Brooks, S.A., Distl, O., Fries, R., Finno, C.J., Gerber, V., Haase, B., Jagannathan, V., Kalbfleisch, T., Leeb, T., Lindgren, G., Lopes, M.S., Mach, N., Machado, A d C., MacLeod, J.N., McCoy, A., Metzger, J., Penedo, C., Polani, S., Rieder, S., Tammen, I., Tetens, J., Thaller, G., Verini-Supplizi, A., Wade, C.M., Wallner, B., Orlande, L., Mickelson, J.R. & McCue, M.E. 2017. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics*. 18: 565.

Schmutz, S.M. & Berryere, T.G. 2007. Genes affecting coat colour and pattern in domestic dogs: a review. *Animal genetics*. 38:6, 539 – 549.

Seabury, C.M., Dowd, S.E., Seabury, P.M., Raudsepp, T., Brightsmith, D.J., Liboriussen, P., Halley, Y., Fisher, C.A., Owens, E., Viswanathan, G. & Tizard,, I.R. 2013. Multi-Platform Draft de novo Genome Assembly and Comparative Analysis for the Scarlet Macaw (*Ara macao*). *PLOS ONE*. 8: e62415.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S. & Seeb, L.W. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*. 11 (Suppl. 1), 1 - 8.

Seth-Smith, D. 1931. Mutant of *A. personata*. Proceedings of the Zoological Society of London 101: 1418.

Simão, F.A. Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31, 19.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M. & Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research*. 19: 1117 - 1123.

Simpson, J.T. & Durbin, R. 2012. Efficient de novo assembly of large genomes using compressed data structures. *Genome Research*. 22: 549 – 556.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. & Ponting, C.P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews*. 15, 121 - 132.

Silva, T. & Kotlar, B. 1988. Breeding lovebirds. Neptune, NJ, United States of America: TFH Publications.

Smit, A.F.A, Hubley, R & Green, P. 2013. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stapley, J., Birkhead, T.R., Burke, T & Slate, J. 2008. A Linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics*. 179: 651 - 667.

Stein, L. 2001. Genome annotation: from sequence to biology. *Nature reviews*. 2:493 - 503.

Stidham, T.A. 2009. A lovebird (Psittaciformes: *Agapornis*) from the Plio-Pleistocene Kromdraai B locality, South Africa. *South African Journal of Science*. 105, 155 - 157.

Stoddard, M.C. & Prum, R.O. 2011. How colourful are birds? Evolution of the avian plumage color gamut. *Behavioral Ecology*. 22: 5, 1042 – 1052.

Stothard, P., Choi J-W., Basu, U., Sumner-Thomson, J.M., Meng, Y., Liao, X. & Moore, S.S. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*. 12:559.

Stradi, R., Pini, E. & Celentano, G. 2001. The chemical structure of the pigments in Ara macao plumage. *Comparative Biochemistry and Physiology Part B.* 130: 57 – 63.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C. 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. *Animal Genetics*. 45, 572 - 575.

Strucken, E.M., Lee, S.H., Song, K.D., Gibson, J.P. & Gondro, G. 2016. How many markers are enough? Factors influencing parentage testing in different livestock populations. *Journal of Animal Breeding and Genetics*. 133, 13 - 23.

Symes, C.T. 2014. Founder populations and the current status of exotic parrots in South Africa. *Ostrich*. 85:3, 235 - 244.

Talenti, A., Nocolazzi, E.L., Chessa, S., Frattini, S., Moretti, R., Coizet, B., Nicoloso, L., Colli, L., Pagnacco, G., Stella, A., Ajmone-Marsan, P., Ptak, G.& Crepaldi, P. 2016. A method for single nucleotide polymorphism selection for parentage assessment in goats. *Journal of Dairy Science*. 99: 3646 - 3653.

Tekalign, W. & Bekele, A. 2011. Status of and threats to the Black-winged Lovebird (*Agapornis taranta*) in Entoto Natural Park and Bole sub-city, Addis Ababa, Ethiopia. *Chinese Birds*. 2: 4, 174 – 182.

The UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Research*. 43: D204 – D212.

Toerien, M.J. 1950. On the change of colour in a rosy-faced lovebird. *Ostrich.* 21: 19 – 21.

Tsudzuki, M. 2008. Mutations of Japanese Quail (*Coturnix japonica*) and recent advances of molecular genetics for this species. *Journal of Poultry Science.* 45: 159 – 179.

Van den Abeele, D. 2016. Lovebirds Compendium, Genus *Agapornis*. The Netherlands: About Pets Publishing.

Van der Jagt, C.J., Chamberlain, A.J., Schnabel, R.D., Hayes, B.J. & Daetwyler, H.D. 2018. Which is the best variant caller for large whole-genome sequencing datasets? Proceedings of the World Congress on Genetics Applied to Livestock Production. 11.128.

van Grouw, H. 2013. What colour is that bird? The causes and recognition of common colour aberrations in birds. *British Birds.* 106: 17 – 29.

Veeckmans, D. 2016. New mutation of Fischer's Lovebird – Agapornis fischeri: Opaline Euwing. Parrots Daily News, 12 April. Available at http://www.parrotsdailynews.com/new-mutation-offischers-lovebird-agapornis-fischeri-opaline-euwing/ [accessed 20 April 2018].

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*. 34, 275 - 305.

Visser, C, van Marle-Köster, E & Friedrich, H. 2011. Parentage verification of South African Angora goats, using microsatellite markers. *South African Journal of Animal Science.* 41: 250 – 255.

vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Parker, H.G., Ostrander, E.A. & Wayne, RK. 2013. Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mammalian Genome.* 24: 80 – 88.

Wajid, B. & Serpedin, E. 2014. Do it yourself guide to genome assembly. *Briefings in functional genomics*. 2014, 1 - 9.

Warburton, L.S. & Perrin, M.R. 2005. Nest-site characteristics and breeding biology of the Black-cheeked Lovebird *Agapornis nigrigenis* in Zambia. *Ostrich.* 76: 3&4, 162 – 174.

Warburton, L.S. & Perrin, M.R. 2006. The Black-cheeked Lovebird (*Agapornis nigrigenis*) as an agricultural pest in Zambia**.** *Emu.* 160**:** 4, 321 – 328.

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., Heger, A., Kong, L., Ponting, C.P., Jarvis, E.D., Mello, C.V., Minx, P., Lovell, P., Velho, T.A.F., Ferris, M., Balakrishnan, C.N., Sinha, S., Blatti, C., London, S.E., Li, Y., Lin, Y-C., George, J., Sweedler, J., Southey, B., Gunaratne, P. & Watson, M. 2010 The genome of a songbird. *Nature*. 464: 757 – 762.

Warren, W.C., Hillier, L.W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K.D., Thibaud-Nissen, F., Schneider, V., Mansour, T.A., Brown, C.T., Zimin, A., Hawken, R., Abrahamsen, M., Pyrkosz, A.B., Morisson, M., Fillin, V., Vignal, A., Chow, W., Howe, K., Fulton, J.E., Miller, M.M., Lovell, P., Mello, C.V., Wirthlin, M., Mason, A.S., Kuo, R., Burt, D.W., Dodgson, J.B. & Cheng, H.H. 2017. A New chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*. 7:1, 109 - 117.

Weinman, L.R., Solomon, J.W. & Rubenstein, D.R. 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*. 15, 502 – 511.

Wirthlin, M., Lima, N.C.B., Guedes, R.L.M., Soares, A.E.R., Almeida, L.G., Caveleiro, N.P.,

de Morais, G.L., Chaves, A.V., Howard, J.T., Teixeira, M. d M., Schneider, P.N., Santos, F.R., Schatz, M.C., Felipe, M.S., Miyaki, C.Y., Aleixo, A., Schneider, M.P.C., Jarvis, E.D. & Mello, C.V. 2018. Parrot genomes and the evolution of heightened longevity and cognition. *Current biology.* 28:24, 4001 - 4008.e7.

Yakovlev, AA, Kudryavtsev, IV, Pigareva, MD. 1975. A mutant Japanese quail (*Coturnix coturnix japonica*) with marble plumage. *Soviet Genetics.* 11: 1225 – 1232.

Yandell, M. & Ence, D. 2012. A beginner's guide to eukaryotic genome annotation. *Nature reviews.* 13: 329 - 341.

Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simao, F.A., Ioannidis, P., Seppey, M., Loetscher, A. & Kriventseva, E.V. 2015. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research.* 45: D744 – D749.

Zerbino, D.R. & Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using *de Bruijn* graphs. *Genome Research*. 18: 821 - 829.

Zhang, G., Li, B., Li, C., Gilbert, M.T.P., Jarvis, E.D., Wong, J. & The Avian Genome Consortium. 2014 (a). Comparative genomic data of the Avian Phylogenomics Project. *GigaScience*. 3, 26.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xiong, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xiong, Y., Wang, G., Wang, B., Wang, J., Fan, Y., da Fonseca, R.R., Alfaro-Núñez, A., Schubert, M., Orlande, L., Mourier, T., Howard, J.T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M.V., Hara, E., Smith, J., Farré, M., Narayan, J., Slavov, G., Romanov, M.N., Borges, R., Machado, J.P., Khan, I., Springer, M.S., Gatesy, J., Hoffmann, F.G., Opazo, J.C., Håstad, O., Sawyer, R.H., Kim, H., Kim, K-W., Kim, H.J., Cho, S., Li, N., Huang, Y., Bruford, M.W., Zhan, X., Dixon, A., Bertelsen, M.F., Derryberry, E., Warren, W., Wilson, R.K., Li, S., Ray, D.A., Green, R.E., O'Brien, S.J., Griffin, D., Johnson, W.E., Hausseler, D., Ryder, OlA., Willerslev, E., Graves, G.R., Alström, P., Fjeldså, J., Mindell, D.P. Edwards, S.V., Braun, E.L., Rahbek, C., Burt, D.W., Houde, P., Zhang, Y., Yang, H., Wang, J., Avian Consortium, Jarvis, E.D., Gilbert, M.T.P., Wang, J. 2014 (b). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 346: 6215, 1311 - 1320.

Zhang, G. 2015. Bird sequencing project takes off. *Nature*. 522, 34.

APPENDIX A: PUBLISHED PAPER I

A review article that focusses on the problems wild lovebird populations face as well as an in-depth discussion on the plumage colour variations found amongst lovebirds was incorporated into this chapter.

This manuscript was featured as the lead article of the journal's issue.

- Published Paper I: **Plumage colour variations in the *Agapornis* genus: a review.** **Henriëtte van der Zwan**, Carina Visser and Rencia van der Sluis.

  Published in: Ostrich (2019) 90:1, 1-10.

Ostrich
Journal of African Ornithology

ISSN: 0030-6525 (Print) 1727-947X (Online) Journal homepage: https://www.tandfonline.com/loi/tost20

Plumage colour variations in the *Agapornis* genus: a review

Henriëtte van der Zwan, Carina Visser & Rencia van der Sluis

To cite this article: Henriëtte van der Zwan, Carina Visser & Rencia van der Sluis (2019) Plumage colour variations in the *Agapornis* genus: a review, Ostrich, 90:1, 1-10

To link to this article: https://doi.org/10.2989/00306525.2018.1540446

Ostrich 2019, 90(1): 1–10
Printed in South Africa — All rights reserved

OSTRICH
ISSN 0030–6525 EISSN 1727-947X
https://doi.org/10.2989/00306525.2018.1540446

Review Paper

Plumage colour variations in the *Agapornis* genus: a review

Henriëtte van der Zwan[1]*, Carina Visser[2] and Rencia van der Sluis[1]

[1] *Centre of Human Metabolomics, North-West University, Potchefstroom, South Africa*
[2] *Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa*
* *Corresponding author, email: agapornis.genome.study@gmail.com*

The genus *Agapornis* consists of nine small African parrot species that are globally well known as pets, but are also found in their native habitat. Illegal trapping, poaching and habitat destruction are the main threats these birds face in the wild. In aviculture, *Agapornis* breeding is highly popular all across the globe. Birds are mainly selected based on their plumage colour variations but very little molecular research has been conducted on this topic. There are 30 known colour variations amongst the nine species and most of these are inherited as Mendelian traits. However, to date none of the genes or polymorphisms linked to these variations have been identified or verified. Due to unethical breeding practices, the need for the development of molecular tests such as identification verification tests or species identity tests is growing. Future research is paramount to ensure the conservation of wild populations as well as aiding breeders in improving breeding strategies.

Variations de couleur du plumage dans le genre *Agapornis*: une critique

Le genre *Agapornis* se compose de neuf petites espèces de perroquets africains qui sont mondialement connus comme animaux de compagnie, mais qui se trouvent également dans leur habitat d'origine. Le piégeage illégal, le braconnage et la destruction de l'habitat sont les principales menaces auxquelles ces oiseaux sont confrontés dans la nature. En aviculture, l'élevage *Agapornis* est très populaire dans le monde entier. Les oiseaux sont principalement sélectionnés en fonction des variations des couleurs de leur plumage, mais très peu de recherches moléculaires ont été menées sur ce sujet. Il existe 30 variations de couleur connues parmi les neuf espèces et la plupart d'entre elles sont héritées des caractères Mendéliens. Cependant, à ce jour, aucun des gènes ou polymorphismes liés à ces variations n'a été identifié ou vérifié. En raison des pratiques de sélection contraires à l'éthique, la nécessité de développer des tests moléculaires tels que des tests de vérification de l'identification ou des tests d'identité d'espèce est de plus en plus importante. Les recherches futures sont primordiales pour assurer la conservation des populations sauvages et aider les éleveurs à améliorer leurs stratégies de sélection.

Keywords: aviculture, lovebirds, molecular breeding tools, Psittacidae

Introduction

Globally, there are 356 extant psittaciform, or parrot, species (Forshaw 1989) of which 20 are native to the African continent and Madagascar (Perrin 2009). Amongst this order, the genus *Agapornis* Selby, 1836 is found and consists of nine species, namely Peach-faced Lovebird *A. roseicollis*, Black-winged Lovebird *A. taranta*, Grey-headed Lovebird *A. canus*, Masked Lovebird *A. personatus*, Nyasa Lovebird *A. lilianae*, Black-collared Lovebird *A. swindernianus*, Fischer's Lovebird *A. fischeri*, Red-faced Lovebird *A. pullarius* and Black-cheeked Lovebird *A. nigrigenis* (Moreau 1948; Dilger 1960; Forshaw 1989). These species are more commonly known as lovebirds and are distributed across Madagascar (Grey-headed Lovebird) and Africa (the remaining eight species) (Dilger 1960; Forshaw 1989; Perrin 2012). Table 1 lists the distribution of each species, size of the birds, known subspecies, common name and popularity of each species in aviculture. In addition to existing in their natural habitat, lovebirds are popular pets and eight of the nine species, the Black-collared Lovebird being the exception, are kept and bred as domestic birds (Hayward 1979; Silva and Kotlar 1988; Forshaw 1989; van Den Abeele 2016).

Lovebirds have been kept as pets since the eighteenth century with reports of Red-faced Lovebirds being bred in Britain in the 1880s (Farner et al. 1982). Hayward (1979) states that lovebirds are popular in aviculture due to the fact that they breed relatively easily in captivity, are hardy, active and have a range of plumage colours. They can also be bred in a small aviary making them ideal birds for today's modern lifestyle. As a consequence of their popularity in aviculture, birds in their natural habitat are being threatened by poaching and trapping for export to pet

Ostrich is co-published by NISC (Pty) Ltd and Informa UK Limited (trading as Taylor & Francis Group)

Published online 01 Mar 2019

markets (Warburton and Perrin 2005, 2006; Mzumara et al. 2016a, 2016b). Plumage colour variations are the main economic factor breeders select for and there are at least 30 colour variations amongst domestic lovebirds that are accepted by shows and auctions (Hayward 1979; van Den Abeele 2016). Rare colour variants are sold for up to 700 times as much as a wildtype colouration bird of the same species in Europe (D van Den Abeele, Ornitho Genetics VZW and MUTAVI, Belgium, pers. comm., 2015) and there are reports of mark-ups of up to 1 300 times in Asia (Diega 2017). All of these colour variants are genetically inherited, but to date the only reports of the mode of inheritance of these traits are breeders' records. Due to many of these colour variants being recessively inherited, it may lead to inbreeding of related birds in order to create a recessive plumage genotype. Given that some lovebird species can inter-mate, hybridisation of different species is also of great concern in aviculture.

Despite their popularity, very little molecular genetic research has been conducted on these species, in particular for plumage colouration inheritance. Currently, the genes and polymorphisms linked to lovebird colouration are not yet known and therefore breeders cannot screen young birds to confirm their colour genotype. There is also no screening test to identify species nor a species-specific parentage panel or identity confirmation test available. These tests could be useful not only in aviculture but also in conservation to monitor natural populations.

This paper reviews the importance of the conservation of the lovebird species in their natural habitat as well as the role these birds play in aviculture. The mechanisms of expression of parrot plumage colour will be discussed as well as the different colour variations that are found in this genus and the importance thereof as an economic driving factor for these species. The lack of genetic screening tests makes genetic progress in aviculture difficult and this paper briefly investigates the different diagnostic tests that could improve lovebird breeding.

The genus *Agapornis*

Lovebirds have been domesticated since the 1880s but surprisingly little molecular genetic research has been conducted on these species. The first taxonomic classification of the species was done by Moreau (1948). The nine extant species were classified into two groups (Group A and B) with two sister taxon groups consisting of the Peach-faced Lovebird and the Black-collared Lovebird, respectively. This classification was revised by Dilger (1960) who classified them only as three different groups based on the presence or absence of the white eye ring, whether they are monomorphic or not, and an intermediate group.

Molecular genetic advances gave researchers the opportunity to study these birds on a more detailed genetic level in the late 1990s. Eberhard (1998) sequenced a portion of the cytochrome *b* (*cytb*) mitochondrial gene of eight lovebird species (with the exception of the Black-collared Lovebird). The molecular evidence from this study confirmed the classification by Dilger (1960) and categorised the species as the eye-ring group, sexually dimorphic group and transitional group. In a later study by Manegold and Podsiadlowski (2014) the *cytb* mitochondrial gene of a Black-collared Lovebird museum sample was additionally analysed and the Black-collared Lovebird was found to be

Table 1: A summary of *Agapornis* species. CAR = Central African Republic, DRC = Democratic Republic of Congo

Species	Distribution	Size (cm)	Subspecies	Common name	Aviculture	References
A. canus	Madagascar	14	*A. c. canus* and *A. c. ablectaneus*	Lavender Head, Grey Head, Madagascar Lovebird	Less common	Hayward (1979); Forshaw (1989); van Den Abeele (2016)
A. taranta	Ethiopia, Eritrea and Djibouti	17	None	Black-winged or Abyssinian Lovebird	Less common	Forshaw (1989); Marez (2003); Tekalign and Bekele (2011); van Den Abeele (2016)
A. pullarius	Across the African equator from Kenya to Guinea over 26 countries	15	*A. p. pullarius* and *A. p. ugandae*	Red-headed Lovebird	Uncommon	Silvia and Kotlar (1988); Forshaw (1989); van Den Abeele (2016); Egwumah and Iboyi (2017)
A. roseicollis	Angola, Namibia, Botswana and South Africa	16	*A. r. roseicollis* and *A. r catumbella*	Peach-faced or Rosy-faced Lovebird	Very common	Forshaw (1989); Ndithia et al. (2007); van Den Abeele (2016)
A. swindernianus	Cameroon, Gabon, Congo, CAR, DRC and Uganda	13	*A. s swinderninus*, *A. s zenkeri* and *A. s emini*	Black-collared Lovebird	Not found	Forshaw (1989); Manegold and Podsiadlowski (2013)
A. personatus	Tanzania	15	None	Masked or Black-masked Lovebird	Very common	Forshaw (1989); Perrin (2009); van Den Abeele (2016)
A. lilianae	Zimbabwe, Zambia and Malawi	14.5	None	Lilian's Lovebird	Very common	Forshaw (1989); Mzumara et al. (2016a, 2016b); van Den Abeele (2016)
A. fischeri	Tanzania	15	None	Fischer's Lovebird	Very common	Forshaw (1989); Mwangomo et al. (2007); van Den Abeele (2016)
A. nigrigenis	Zambia	14	None	Black-cheeked Lovebird	Very common	Forshaw (1989); Warburton and Perrin (2005, 2006); van Den Abeele (2016)

a sister clade to all other lovebird species. The Grey-headed Lovebird, Red-faced Lovebird, Black-collared Lovebird and Peach-faced Lovebird all have known subspecies (Table 1). Some of the subspecies are poorly differentiated and researched (Forshaw 1989; van Den Abeele 2016).

Karyotyping of the Peach-faced Lovebird genome by Nanda et al. (2007) revealed that this species has a diploid number of 48 macro-chromosomes. In the same study the Budgerigar *Melopsittacus undulatus* was found to have $2n = 62$ macro-chromosomes. The lower number of lovebird chromosomes might indicate a fusion of ancestral micro-chromosomes (Nanda et al. 2007). The budgerigar had a marginally larger genome size of 1.2 Gb (Ganapathy et al. 2014) compared with the Peach-faced Lovebird genome of 1.1 Gb (van der Zwan et al. 2018).

Crossbreeding between some of the species is possible and can produce viable hybrids (McCarthy 2006; van Den Abeele 2016). The species' habitat ranges do not overlap and therefore interbreeding is rarely found in the wild. The only habitat overlap is between the Red-headed Lovebird and Black-collared Lovebird (Moreau 1948; Dilger 1960; Forshaw 1989) but these species' lifestyles are so different that, even though not well researched, it is doubted that interbreeding occurs. In aviculture, unfortunately, it is not uncommon to see hybridisation of species. Hybrids are not generally accepted at auctions and shows (van Den Abeele 2016) and crosses between species should not be encouraged in aviculture. McCarthy (2006) lists the possible observed crosses between lovebird species that could produce viable young and a summary is given in Table 2 (viable young born is depicted with an 'x').

Viable offspring hatching from a mating between a Grey-headed Lovebird male and Budgerigar female, Black-cheeked Lovebird male and Budgerigar female, Masked Lovebird male and Budgerigar female, and a Peach-faced Lovebird male and Budgerigar female (McCarthy 2006) has also been reported. It is well known that some plumage colour variations were introduced to other lovebird species by intentional crossing of the species that can intermate (van Den Abeele 2016).

Ecology and threats to natural populations

The behaviour and habitat of the nine species varies greatly. Black-collared Lovebirds are exclusively forest canopy dwellers, Nyasa Lovebirds prefer riverbeds and mopane *Colophospermum mopane* woodlands, the Black-winged Lovebird lives at high altitudes of 1 600–3 800 m, Red-faced Lovebirds are lowland tropical birds, the Peach-faced Lovebird is found in semi-arid environments, Fischer's Lovebirds and Masked Lovebirds are lowland savanna and woodland dwellers, Grey-headed Lovebirds prefer lightly wooded habitats and the Black-cheeked Lovebird can be found in mopane woodlands (Forshaw 1989; Perrin 2009; van Den Abeele 2016). The main diet of most of the lovebird species consists of grass seeds, but Black-collared Lovebirds feed on ripened figs (Forshaw 1989; Perrin 2009). Moreau (1948) notes that Black-collared Lovebirds cannot survive in captivity if not fed figs, which could explain why these birds are not kept as pets.

The International Union for Conservation of Nature (IUCN) lists the Black-cheeked Lovebird as Vulnerable, Fischer's Lovebirds and Nyasa Lovebirds as Near Threatened and the other species as Least Concern. All species except the Peach-faced Lovebird are classified by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) as CITES Appendix II species, indicating that they are not threatened with extinction, but trade must be controlled to prevent their extinction. Several studies have been conducted on the natural lovebird populations and the threats these birds are faced with. In conclusion, Tekalign and Bekele (2011) (on the Black-winged Lovebird), Mzumara et al. (2016a, 2016b) (on Nyasa Lovebirds), Egwumah and Iboyi (2017) (on Red-faced Lovebirds) and Perrin (2012) (all species) found that reduction of their natural habitat due to agricultural activity and trapping or poaching are the main threats to these species. In addition, Mzumara et al. (2016a) also found that poisoning of birds in Malawi is a common practice as these birds destroy crops.

Worldwide, there is keen interest in lovebird breeding. Membership of the world's largest breeding association – the Belgium Agapornis Breeders Association (BVA-International) – is approximately 5 000 members (D van Den Abeele, pers. comm.) and other societies across Europe, Africa, Australia, the USA, South America and Asia (van Den Abeele 2016) exist with various member numbers. In a 2017 poll conducted by Ornitho Genetics VZW (https://www.ornitho-genetics.info/), the most popular lovebird species bred in captivity was the Fischer's Lovebird, followed by Masked Lovebirds and Peach-faced Lovebirds. The popularity of a species in aviculture could drastically influence their numbers in the natural habitat as traders will trap these birds in order to sell them. Perrin (2012) reports that Peach-faced Lovebirds, Masked Lovebirds and Fischer's Lovebirds are all heavily traded, whereas Nyasa Lovebirds are traded moderately and Black-cheeked Lovebirds to a lesser extent. Trade and trappings are banned in several African countries, which should have an effect on the natural populations' growth.

Table 2: Possible viable matings amongst domestic *Agapornis* species. Viable young born is depicted with X

	A. canus	*A. fischeri*	*A. nigrigenis*	*A. lilianae*	*A. personatus*	*A. roseicollis*	*A. pullarius*	*A. taranta*
A. canus		X	X	X	X	X	X	X
A. fischeri	X		X	X	X	X		X
A. nigrigenis	X	X		X	X	X		
A. lilianae	X	X	X		X	X		
A. personatus	X	X	X	X		X	X	X
A. roseicollis	X	X	X	X	X		X	
A. pullarius	X				X	X		X
A. taranta	X	X			X			

Aviculturists have the responsibility to breed with purebred, healthy birds to ensure that if the wild populations go extinct, these species will still survive in aviculture. With the use of molecular diagnostic tests, breeders will be able to select birds with recessive colour variations without compromising the biodiversity of the species. Unrelated birds that are heterozygous for a recessive colour trait could be identified and mated, rather than mating close relatives that are suspected heterozygotes. Breeders should be made aware of the importance of conserving wild populations as well as ethical breeding practices. This could be attained by promoting the importance of not selecting breeding pairs based on one trait (colour) only, but to keep biodiversity and general genetic fitness in mind.

Lovebird plumage colouration

As mentioned in the Introduction, the main economic factor breeders select for is plumage colouration and many of these are inherited as autosomal recessive traits (Table 3). This implies that an offspring with wildtype colouration has a certain chance (depending on the parents' genotypes) of being a heterozygote (thus a carrier or split) of this trait. Breeders sell possible heterozygotes at a premium despite the fact that the buyer has neither proof that the bird is an offspring of the colour variant nor that it is indeed a heterozygote for the colour variation. This could lead to fraudulent transactions and emphasises the need for more research in the field of plumage colour genetics and parentage verification testing for lovebirds.

Lovebird plumage colour and biological mechanisms controlling parrot colouration

The majority of the lovebird species has at least one colour variation that is found infrequently in wild populations and, due to selective breeding, frequently in domestic populations. The main colour of all wildtype lovebirds is green with red, yellow, pink, black or grey on the head, chest and neck, and some blue in the tail feathers (Hayward 1979; Silva and Kotlar 1988; Forshaw 1989; van Den Abeele 2016), as seen in the photographs of the wildtype colouration of the eight domestic lovebird species in Figure 1. In Figures 2a–h some of the colour variants of the different species are shown with the wildtype variant of that species. Reports of wild lovebirds with colours other than the wildtype colouration are scarce but Warburton and Perrin (2005) observed one yellow and three light-green Black-cheeked Lovebirds in the mid Machile River region between May 1999 and

Table 3: Plumage colour variations and their inheritance patterns found amongst *Agapornis* species. NSL# = Non-sexed linked ino, SL† = sexed linked ino, AID = autosomal incomplete dominance, AR = autosomal recessive, SLR = sex-linked recessive, AD = autosomal dominant. X depicts the mutations found amongst each species

	A. roseicollis	*A. canus*	*A. taranta*	*A. pullarius*	*A. personatus*	*A. fischeri*	*A. lilianae*	*A. nigrigenis*	Inheritance pattern
Dark factor	X		X		X	X	X		AID
Blue					X	X	X	X	AR
Aqua	X					X			AR
Turquoise	X		X						AR
Orange face	X								AR
Pale headed	X								AID
Opaline	X					X			SLR
Ino (NSL#)					X	X	X	X	AR
Ino (SL†)	X								SLR
Pastel					X				AR
Dark eyed clear						X			AR
Pallid	X								SLR
Cinnamon	X								SLR
Pale	X					X			SLR
Marbled	X								AR
Dilute	X		X				X	X	AR
Bronze fallow	X		X			X			AR
Pale fallow	X		X			X			AR
Dun fallow					X				AR
Dominant pied	X					X			AD
Recessive pied	X					X			AR
Mottle						X			Unclear inheritance
Dominant edged						X			AID
Misty	X		X			X		X	AID
Slaty						X			AID
Violet	X				X				AID
Euwing						X			AID
Faded	X					X			AR
Crested	X					X			Incomplete dominance
Jade	X								AR

May 2000. Van Den Abeele (2016) identified a wild-caught double factor green Peach-faced Lovebird in Namibia and in the 1931 *Proceedings of the Zoological Society of London* a wild-caught Masked Lovebird male was the first blue Masked Lovebird documented (Seth-Smith 1931). Some of these variations have been known to breeders for almost a century. Toerien (1950) reports a lutino Peach-faced Lovebird bred from a wild-caught population from Namibia, whereas Delacour (1942) reports the same variation in Fischer's Lovebirds and Moreau (1948) in Nyasa Lovebirds.

In most avian species, plumage colouration is affected by genetically controlled dietary intake of carotenoids (Sage 1962; Guay et al. 2012). However, in the case of parrots, diet has little or no influence on colouration as it is mainly due to genetic variation (Hill and McGraw 2006). Black, brown and red-brown plumage, beak, claw and eye colours in all bird species are caused by the pigment melanin, whereas red, yellow, pink and orange colours found in parrot plumage are caused by psittacofulvin pigments (Mundy 2006; Hill and McGraw 2006). Parrots are the only group of organisms known to synthesise psittacofulvin pigments in their plumage cells (Brush 1990; Stradi et al. 2001; McGraw and Nogare 2005; Hill and McGraw 2006). Pigments (melanin and psittacofulvins) and structural changes in the feather barbs are the two mechanisms that produce the various parrot colours (Stoddard and Prum 2011). Physical interactions of light waves with β-keratin rods in the feather barbs produces purple, blue and green colours (Dyck 1971a, 1971b; Prum et al. 1994; Prum 2006). The two mechanisms work together as follows: pigments (melanins and psittacofulvins) absorb and reflect light selectively (D'Alba et al. 2012) and due to the structural changes in the feather barbs the light is reflected irregularly (Dyck 1971a, 1971b; Hill and McGraw 2006; Berg and Bennett 2010). The barb of a green feather, for example, contains a cortex filled with yellow psittacofulvin pigment. Light is partially absorbed by the eumelanin pigment in the feather, and blue light is reflected back through the air vacuoles. It then passes through the yellow psittacofulvin pigment and through coherent scattering of the quasi-ordered β-keratin rods, the light (and ultimately the feather) appears green (Dyck 1971a, 1971b). Dark green feathers reflect about half the light compared with that of lighter green feathers (Dyck 1971a, 1971b). Mundy (2006) stresses the fact that these two mechanisms (pigment and structural changes) are most probably under independent genetic control but that no candidate genes have been associated with structural colouration control.

Lovebird plumage variations

Buckley (1987) defines a plumage colour polymorphism as the presence of a plumage colour aberration in an interbreeding population of individuals of the same sex and age. There are at least 30 known naturally occurring plumage colour variations amongst six of the domestic lovebird species (van Den Abeele 2016). The inheritance

Figure 1: Eight domestic *Agapornis* species with wildtype colouration. (a) *A. canus* (male left, female right), (b) *A. taranta* (male left, female right), (c) *A. pullarius* (male left, female right), (d) *A. roseicollis*, (e) *A. personatus*, (f) *A. fischeri*, (g) *A. nigrigenis* and (h) *A. lilianae* (all photographs: Mr Dirk van Den Abeele)

Figure 2: Wildtype colouration and some colour variations of four *Agapornis* species. (a) *A. fischeri*: Wildtype, (b) *A. fischeri*: Turquoise, (c) *A. personatus*: Wildtype, (d) *A. personatus*: blue (left blue, middle DD blue [homozygous dark factor], right D blue [heterozygous dark factor]), (e) *A. lilianae* Wildtype, (f) *A. lilianae* Non-sex linked ino green, (g) *A. roseicollis*: Wildtype, (h) *A. roseicollis*: Opaline d blue (all photographs: Mr Dirk van Den Abeele)

patterns of these variations have mainly been determined by test matings performed by breeders, and very little scientific literature is available on this subject. MUTAVI, a parrot breeding research and advice group located in The Netherlands (http://www.mutavi.info) conducted research trials on the structures of lovebird feathers of various colourations. In Table 3, a summary of the different plumage colour variations per species, plus their inheritance patterns, is given. Given that none of the polymorphisms linked to these colour variations have been identified, the colour patterns will be referred to as 'plumage colour variations' and not 'plumage colour polymorphisms'.

There are four main categories of colour variations in lovebirds and a discussion with examples thereof follows below. These include variations due to psittacofulvin reduction or modification, variations due to a change in melanin, change in feather structure and changes in eumelanin. It is important to note that many of these plumage colour variations can be combined that will result in a range of plumage colours (Hayward 1979; van Den Abeele 2016). Berg and Bennett (2010) as well as Mundy (2018) highlights the lack of research on the genetic control of aberrant colours in caged parrots.

Aberrant plumage colouration due to psittacofulvin reduction or modification

Some of the most dramatic colour variations amongst parrots is caused by the reduction or modification of psittacofulvin

within the feather cortex. In the 1932 *Proceedings of the Zoological Society of London* a wild-caught Masked Lovebird male was described as 'the yellow coloring was absent, the areas of the body normally yellow, being whitish and those normally green, blue.' Cooke et al. (2017) reported that since the nineteenth century Budgerigars bred in captivity have displayed a loss of psittacofulvin pigmentation causing birds to have blue, instead of the wildtype green, plumage colouration. Yellow pigment in the cortex of the feather is normally combined with blue light and results in visibly green feathers. Cooke et al. (2017) found that due to a T→C substitution mutation in the *MuPKS* gene in Budgerigars, no yellow psittacofulvin pigment is produced and therefore blue light is reflected so that the feather appears blue. Other areas of the body that would normally be red, yellow or orange due to psittacofulvins are white due to the absence of the psittacofulvin pigment (Prum 2006; van Den Abeele 2016). The same phenotype is observed amongst lovebirds and it is also inherited as an autosomal recessive trait, as in Budgerigars. Research done at MUTAVI (unpublished data, D van Den Abeele, pers. comm.) has found that no yellow psittacofulvin pigment was present in blue lovebirds compared with pigment being present in wildtype green birds. More research on the molecular basis of this gene in lovebirds is needed before conclusions can be made for this genus.

In contrast to the Blue phenotype where no psittacofulvin is produced, birds with Aqua and Turquoise phenotypes produce a reduced amount of psittacofulvin. There are no reports of studies where the actual pigment is measured, but visibly about 50% of the normal amount of psittacofulvin is produced and only about 40% of the normal levels of psittacofulvin is visible in birds with the Turquoise variation (van Den Abeele 2016). It remains to be confirmed if a *MuPKS* polymorphism is causative of these phenotypes.

Orange Face and Pale Headed are two popular phenotypes in lovebird aviculture and are only found amongst Peach-faced Lovebirds (van Den Abeele 2016). The mechanisms behind these phenotypes are poorly understood (D van Den Abeele, pers. comm.), but for Orange Face individuals red plumage is changed to orange, whereas in Pale Headed individuals the psittacofulvin is changed from red to pink. Green plumage remains unchanged for both phenotypes. Research by McGraw and Nogare (2005) indicated that the red, orange and pink colourations are all due to psittacofulvin pigment in different concentrations and therefore these colour variations could be caused by a change in psittacofulvin pigment concentration, but this remains to be confirmed.

Colour variations due to a change in melanin

Mundy (2006) and McGraw (2006) noted that most colour variations are caused due to changes in melanin distribution. However, little to no molecular research has been conducted on parrot species regarding the genetic basis of these variations. Most of the common types of bird plumage colour variations can be greatly attributed to a lack of melanin in the feathers. Examples of plumage colouration variations due to a change in melanin in lovebirds are briefly given in Table 4.

Table 4: Colour variations due to a change in melanin in *Agapornis* species

Aberration	Phenotype	Reference
Opaline	Rearrangement of colour pigments causing unusual colour patterns	Hume and van Grouw (2014)
Ino sex-linked (ino-SL)	Green > Yellow Blue > White Dark red eye Pale pink feet	Gunnarsson et al. (2007); van Den Abeele (2016)
Ino non-sex-linked (ino-NSL)	Green > Yellow Blue > White Bright red eye	van Den Abeele (2016)
Pastel	Paler green colour	van Grouw (2013); van Den Abeele (2016)
Dec	Primarily yellow with a green hue	van Den Abeele (2016)
Pallid	Hatch with red eyes turn darker	van Den Abeele (2016)
Pale	Paler than wildtype	van Den Abeele (2016)
Dilute	Very light green/yellow colour	van Den Abeele (2016)
Misty	Heterozygotes: dull colour; homozygotes: resemble double factor green variant	van Den Abeele (2016)
Bronze-, Pale- and Dun-fallow	Some eumelanin still present, has red eyes	van Den Abeele (2016)
Cinnamon	Black feathers are brown, paler green, lighter blue rump	Hayward (1979); Tsudzuki (2008)
Dominant and recessive Pied (piebald)	Green > Yellow Blue > White Not expressed in all feathers	Hayward (1979); Guay et al. (2012); van Grouw (2013); van Den Abeele (2016)
Marble	Feather tips have a darker edge	Yakovlev et al. (1975); van Den Abeele (2016)
Dominant edged	Heterozygotes: clear darker edges on wingtips; homozygotes: yellow with light grey flight feathers and a green hue on wings	van Den Abeele (2016)

Colour variations due to a change in feather structure

The 'dark factor' phenotype causes the spongy zone of the feather to decrease in size. Given this reduction only about half the light is reflected throughout the visible spectrum compared with that of light green feathers (Dyck 1971a, 1971b). The inheritance of this phenotype is incomplete dominance and three phenotypes are observed. Heterozygote birds (also called 'single dark factor') are slightly darker (a khaki green) than the wildtype and birds that are homozygous for the mutation ('double dark factor') are found to be an olive-green colour (Hayward 1979; van Den Abeele 2016; MUTAVI unpublished data). This phenotype is observed in five of the lovebird species as seen in Table 3. It is also found amongst Budgerigars resulting in the same colours (Harris 1979).

Research projects conducted at MUTAVI discovered that lovebirds with the Slaty phenotype had a change in the keratin in the feathers from a horn colour to a glassy see-through colour (D van Den Abeele pers. comm.) and plumage colour changed to a steel-blue colour. A similar phenotype, called Slate, is commonly found amongst Budgerigars (Elliott 2013) and is caused by changes in the feather structure and inherited as a sex-linked recessive trait. Through test matings it was determined that this phenotype is inherited as an incomplete dominant trait. However, there is visually no difference between a heterozygote and a homozygote for the mutation and therefore genotypes cannot be determined based on phenotype alone (van Den Abeele 2016). Further research is necessary to identify the gene and polymorphism linked to these phenotypes and confirmation whether the Slaty and Slate phenotypes are caused by the same polymorphism.

The Violet phenotype develops due to a structural change in the spongy zone of the feather. It is, however, only visible when it is combined with the blue phenotype. If psittacofulvin is absent in the cortex (thus a blue bird), combined with the structural change in the spongy zone, the bird appears violet. The same phenotype is found in Budgerigars with the same underlying inheritance pattern (Harris 1979; van Den Abeele 2016). There is a difference in appearance between a heterozygote and homozygote for the mutation in both of the species.

Colouration due to changes in eumelanin

In 2017 the Jade phenotype was officially accepted as a plumage colour aberration of the Peach-faced Lovebird, and is inherited as an autosomal recessive trait. Interestingly the breeders of these birds observed that females are lighter than males. This phenotype has not been described in the literature and all observations are from personal communication with Mr Dirk van Den Abeele. Birds with this variation appears to have a visible reduction of about half of the eumelanin in their wing feathers and more so over their body, resulting in a yellow colour with a darker hue. There is a greater eumelanin reduction in the core of the wing feathers, which results in a slight edge on the feathers. The bird's rump is much lighter and the legs are of normal colouration, but the nails are darker. The mask of these birds is of normal colouration.

The Euwing pattern is a colour variation only found amongst Fischer's Lovebirds causing darker wings and an overall different body colour. The cause of the phenotype is still unclear but is inherited as a dominant trait with incomplete penetrance (van Den Abeele 2016). Veeckmans (2016) notes that the Euwing mutation 'dulls' the black eumelanin pigment in the body feathers and on the mantle in heterozygote birds. In homozygous affected birds, the effect is much more prominent with additional eumelanin found in the wing covert feathers, making the bird appear darker. By combining a bird with both the Euwing mutation as well as the Opaline mutation, breeders can achieve spectacular colourations.

Parentage verification

Applying molecular techniques to confirm parentage has been well documented for various domesticated animal species, including cattle (Matukumalli et al. 2009), goats (Visser et al. 2011) and dogs (DeNise et al. 2003). Marker panels for parentage verification in birds are also found, but these panels are applied more in conservation than breeding (Jan and Fumagalli 2016; Coetzer et al. 2017). Most of these studies used microsatellite markers, but in recent years the focus has shifted towards the use of single nucleotide polymorphisms (SNPs) (see Weinman et al. 2015 for a comparison). Identifying incorrect pedigrees are of importance in livestock populations as it adversely influences genetic gains (Israel and Weller 2000). Although lovebird species are not under directional selection for increased genetic gain, incorrect pedigrees will still negatively affect the populations' general fitness. Despite breeders' efforts to separate breeding pairs and recording the offspring, misidentification of parents still occurs. Due to the fact that many of the colour variations are inherited as autosomal recessive traits, breeders may deliberately sell a young as an offspring of a bird with the colour variation knowing there is no way of confirming either parentage or the colour genotype. This creates a serious loophole that could be exploited by unethical breeders. Currently, there is no species- or genus-specific parentage and individual identification panel available for lovebirds. The development of such a panel could benefit breeders, buyers and conservationists in not only conserving these species but also ensuring ethical breeding practices.

Species identification

Rheindt and Edwards (2011) report that hybridisation, as well as the resulting genetic introgression, is not uncommon in birds. In aviculture the crossing of two species might result in the introduction of a novel colour variation in a different species. Given that birds from some of the different lovebird species can inter-mate (see Table 2), many of the variations were transferred through genetic introgression from one species to another. For example, the 'dark factor' phenotype was first observed amongst Peach-faced Lovebirds in 1968 and in the 1980s the first dark factor Masked Lovebird was identified. From there the phenotype was transferred to Fischer's Lovebird, Nyasa Lovebird and Black-cheeked Lovebird by interbreeding of the species (van Den Abeele 2016). Responsible aviculturists should take care not to interbreed species just to develop 'new'

colours in a species or to develop a 'new lovebird species'. Hybrids are not generally accepted by breeding societies, but unknowing buyers might purchase a hybrid and breed further with this bird, leading to an even larger hybrid population. In addition, conservationists and custom officials can benefit from a species identification test to confirm species from a tissue sample. Genetic testing to infer if a species or an individual is a hybrid between two species has been applied frequently in aquaculture (Rasmussen and Morrissey 2008; Hohenlohe et al. 2011). VonHoldt et al. (2013) reports the application of SNPs to distinguish between dog and wolf hybrids, and Koutsogiannouli et al. (2010) applied restriction fragment length polymorphisms to identify hybrids between wild and domestic pigs. The development of a similar panel of SNPs could aid breeders, breeding societies, buyers, conservationists and custom officials in identifying from which species an individual or a tissue sample originated and whether the individual is a purebred or hybrid bird.

Conclusion

The popularity of lovebirds in aviculture coupled with their reduced habitat and poaching has placed the natural populations of these birds under strain. It is the aviculturists' duty to breed the birds responsibly to ensure the survival of these species should the wild population become extinct. More research is needed in the field of avian genetics to determine the causative genes and mutations linked to colour variations and to develop genetic screening tests to determine the genotype, species and pedigree of the bird.

Acknowledgements — We would like to thank Mr Dirk van Den Abeele for sharing his insight, years of knowledge and photographs with us. The study was funded by the Technology and Human Resources for Industry Programme (THRIP), which forms part of the Department of Trade and Industry of the South African government project ID number TP13082831254, Lumegen Laboratories as well as the Technology Innovation Agency, SABDI grant (407/01 SABDI 16/1016).

ORCID

Henriëtte van der Zwan https://orcid.org/0000-0001-5288-7609
Carina Visser https://orcid.org/0000-0002-3335-4679
Rencia van der Sluis https://orcid.org/0000-0003-3463-1700

References

Berg ML, Bennett ATD. 2010. The evolution of plumage colouration in parrots: a review. *Emu* 110: 10–20.
Brush AH. 1990. Metabolism of carotenoid pigments in birds. *FASEB Journal* 4: 2969–2977.
Buckley PA. 1987. *Avian genetics: a population and ecological approach*. London: Academic Press.
Coetzer WG, Downs CT, Perrin MR, Willows-Munro S. 2017. Testing of microsatellite multiplexes for individual identification of Cape Parrots (*Poicephalus robustus*): paternity testing and monitoring trade *PeerJ* 5: e2900.
Cooke TF, Fischer CR, Wu P, Khosla C, Chuong C-M, Bustamante CD. 2017. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell* 171: 427–439.
D'Alba L, Kieffer L, Shawkey MD. 2012. Relative contributions of pigments and biophotonic nanostructures to natural color production: a case study in budgerigar (*Melopsittacus undulatus*) feathers. *Journal of Experimental Biology* 215: 1272–1277.
Delacour J. 1942. Bird breeding in France. *Avicultural Magazine, 5th Series* 7: 28.
DeNise S, Johnston E, Halverson J, Marshall K, Rosenfeld D, McKenna S, Sharp T, Edwards J. 2003. Power of exclusion for parentage verification and probability of match for identity in American kennel club breeds using 17 canine microsatellite markers. *Animal Genetics* 35: 14–17.
Diega A. 2017. Bird flu to increase foreign demand for African lovebirds. *Business Mirror*, 21 May. Available at https://businessmirror.com.ph/bird-flu-to-increase-foreign-demand-for-african-lovebirds/ [accessed 18 April 2018].
Dilger W. 1960. The comparative ethology of the African parrot genus *Agapornis*. *Zeitschrift für Tierpsychologie* 17: 649–685.
Dyck J. 1971a. Structure and colour-production of the blue barbs of *Agapornis roseicollis* and *Cotinga maynana*. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 115: 17–29.
Dyck J. 1971b. Structure and spectral reflectance of green and blue feathers of the rose-faced lovebird (*Agapornis roseicollis*). *Bilogiske Skrifter* 18(2). Copenhagen: Kongelige Danske Videnskabernes Selskab.
Eberhard JR. 1998. Evolution of nest-building behavior in *Agapornis* parrots. *The Auk* 115: 455–464.
Egwumah FA, Iboyi MO. 2017. Paramount factors influencing the breeding performance of lovebird *Agapornis pullaria*. *International Journal of Life Sciences and Technology* 10: 69–74.
Elliot FS. 2013. *Budgerigar matings and colour expectations*. UK: Read Books.
Farner DS, King JR, Parkes KC (eds). 1982. *Avian biology*, vol. 6. New York: Academic Press.
Forshaw J. 1989. *Parrots of the world*. Princeton: Princeton University Press.
Ganaparthy G, Howard JT, Ward JM, Li J, Li B, Li Y, Xiong Y, Zhang Y, Zhou S, Schwartz DC, Schatz M, Aboukhalil R, Fedrigo O, Bukovnik L, Wang T, Wray G, Rasolonjatovo I, Winer R, Knight JR, Koren S, Warren WC, Zhang G, Phillippy AM, Jarvis ED. 2014. High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience* 3: 11.
Guay P-J, Potvin DA, Robinson RW. 2012. Aberrations in plumage coloration in birds. *Australian Field Ornithology* 29: 23–30.
Gunnarsson U, Hellström AR, Tixier-Boichard M, Minvielle F, Bed'hom B, Ito S, Jensen P, Rattink A, Vereijken A, Andersson L. 2007. Mutations in *SLC45A2* cause plumage color variation in chicken and Japanese quail. *Genetics* 17: 857–577.
Harris RL. 1979. Genetics and the budgerigar. *AFA Watchbird Journal of the American Federation of Aviculture* 6: 2.
Hayward J. 1979. *Lovebirds and their colour mutations*. Poole: Blandford Press.
Hill GE, McGraw KJ (eds). 2006. *Bird coloration, vol. 1: mechanisms and measurements*. Cambridge, MA: Harvard University Press.
Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* 11 (Suppl. 1): 117–122.
Hume JP, van Grouw H. 2014. Colour aberrations in extinct and endangered birds. *Bulletin of the British Ornithologists' Club* 134: 168–193.
Israel C, Weller JI. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *Journal of Dairy Science* 83: 181–187.
Jan C, Fumagalli L. 2016. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). *PeerJ* 4: e2416.

Koutsogiannouli EA, Moutou KA, Sarafidou T, Stamatis C, Mamuris Z. 2010. Detection of hybrids between wild boars (*Sus scrofa scrofa*) and domestic pigs (*Sus scrofa* f. *domestica*) in Greece, using the PCR-RFLP method on melanocortin-1 receptor (MC1R) mutations. *Mammalian Biology* 75: 69–73.

Manegold A, Podsiadlowski L. 2014. On the systematic position of the Black-collared Lovebird *Agapornis swindernianus* (*Agapornithinae*, Psittaciformes). *Journal of Ornithology* 155: 581–589.

Marez C. 2003. The Abyssinian Lovebird *Agapornis taranta*. *AFA Watchbird* 30(3): 20–23.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connoll J, Moore SS, Smith TPL, Sonstegard TS, van Tassell CP. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLOS ONE* 4: e5350.

McCarthy E. 2006. *Handbook of avian hybrids of the world*. Oxford: Oxford University Press.

McGraw KJ. 2006. Mechanics of uncommon colors: pterins, porphyrins, and psittacofulvins. In: Hill GE, McGraw KJ (eds), *Bird coloration, vol. 1: Mechanisms and measurements*. Cambridge, MA: Harvard University Press. pp 354–398.

McGraw KJ, Nogare MC. 2005. Distribution of unique red feather pigments in parrots. *Biology Letters* 1: 38–43.

Moreau R. 1948. Aspects of evolution in the parrot genus *Agapornis*. *Ibis* 90: 206–239.

Mundy NI. 2006. A window on the genetics of evolution: *MC1R* and plumage coloration in birds. *Proceedings of the Royal Society B* 272: 1633–1640.

Mundy NI. 2018. Colouration genetics: pretty polymorphic parrots. *Current Biology* 26: R113–R114.

Mwangomo EA, Hardesty LH, Sinclair ARE, Mduma SAR, Metzger KL. 2007. Habitat selection, diet and interspecific associations of the rufous-tailed weaver and Fischer's lovebird. *African Journal of Ecology* 46: 267–275.

Mzumara TI, Perrin MR, Downs CT. 2016a. The drinking habits of the Lilian's lovebird and incidents of poisoning at waterholes. *African Journal of Ecology* 54: 289–298.

Mzumara TI, Perrin MR, Downs CT. 2016b. Comparative performance of four survey methods for assessing Lilian's Lovebird abundance in Liwonde National Park, Malawi. *Ostrich* 8: 247–253.

Nanda I, Karl EK, Griffin DK, Schartl M, Schmid M. 2007. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenetic and Genome Research* 117: 43–53.

Ndithia H, Perrin MR, Waltert M. 2007. Breeding biology and nest site characteristics of the Rosy-faced Lovebird *Agapornis roseicollis* in Namibia. *Ostrich* 78: 13–20.

Perrin MR. 2009. Niche separation in African parrots. In: Harebottle DM, Craig AJFK, Anderson MD, Rakotomanana H, Muchai M (eds), *Proceedings of the 12th Pan-African Ornithological Congress, 2008*. Cape Town: Animal Demography Unit. pp 29–37.

Perrin M. 2012. *Parrots of Africa, Madagascar and the Mascarene islands: biology, ecology and conservation*. Johannesburg: Wits University Press.

Prum RO. 2006. Anatomy, physics, and evolution of structural colours. In: Hill GE, McGraw KJ (eds), *Bird coloration, vol. 1: Mechanisms and measurements*. Cambridge, MA: Harvard University Press. pp 295–353.

Prum RO, Morrison RL, ten Eyck GR. 1994. Structural color production by constructive reflection from ordered collagen arrays in a bird (*Philepitta castanea*: Eurylaimidae). *Journal of Morphology* 222: 61–72.

Rasmussen RS, Morrissey MT. 2008. DNA-based methods for the identification of commercial fish and seafood species. *Comprehensive Reviews in Food Science and Food Safety* 7: 280–295.

Rheindt FE, Edwards SV. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *The Auk* 128: 620–632.

Sage BL. 1962. Albinism and melanism in birds. *British Birds* 55: 6.

Seth-Smith D. 1931. Mutant of *A. personata*. *Proceedings of the Zoological Society of London* 101: 1418.

Silva T, Kotlar B. 1988. *Breeding lovebirds*. Neptune, NJ: TFH Publications.

Stoddard MC, Prum RO. 2011. How colorful are birds? Evolution of the avian plumage color gamut. *Behavioral Ecology* 22: 1042–1052.

Stradi R, Pini E, Celentano G. 2001. The chemical structure of the pigments in *Ara macao* plumage. *Comparative Biochemistry and Physiology Part B* 130: 57–63.

Tekalign W, Bekele A. 2011. Status of and threats to the Black-winged Lovebird (*Agapornis taranta*) in Entoto Natural Park and Bole sub-city, Addis Ababa, Ethiopia. *Chinese Birds* 2: 174–182.

Toerien MJ. 1950. On the change of colour in a rosy-faced lovebird. *Ostrich* 21: 19–21.

Tsudzuki M. 2008. Mutations of Japanese Quail (*Coturnix japonica*) and recent advances of molecular genetics for this species. *Journal of Poultry Science* 45: 159–179.

van Den Abeele D. 2016. *Lovebirds compendium, genus* Agapornis*: species, breeding, genetics, mutations*. The Netherlands: About Pets.

van der Zwan H, van der Westhuizen F, Visser C, van der Sluis R. 2018. Draft *de novo* genome sequence of *Agapornis roseicollis* for application in avian breeding. *Animal Biotechnology* 29: 241–246.

van Grouw H. 2013. What colour is that bird? The causes and recognition of common colour aberrations in birds. *British Birds* 106: 17–29.

Veeckmans D. 2016. New mutation of Fischer's Lovebird – *Agapornis fischeri:* Opaline Euwing. *Parrots Daily News*, 12 April. Available at http://www.parrotsdailynews.com/new-mutation-of-fischers-lovebird-agapornis-fischeri-opaline-euwing/ [accessed 20 April 2018].

Visser C, van Marle-Köster E, Friedrich H. 2011. Parentage verification of South African Angora goats, using microsatellite markers. *South African Journal of Animal Science* 41: 250–255.

vonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK. 2013. Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mammalian Genome* 24: 80–88.

Warburton LS, Perrin MR. 2005. Nest-site characteristics and breeding biology of the Black-cheeked Lovebird *Agapornis nigrigenis* in Zambia. *Ostrich* 76: 162–174.

Warburton LS, Perrin MR. 2006. The Black-cheeked Lovebird (*Agapornis nigrigenis*) as an agricultural pest in Zambia. *Emu* 160: 321–328.

Weinman LR, Solomon JW, Rubenstein DR. 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources* 15: 502–511.

Yakovlev AA, Kudryavtsev IV, Pigareva MD. 1975. A mutant Japanese quail (*Coturnix coturnix japonica*) with marble plumage. *Soviet Genetics* 11: 1225–1232.

Received 14 May 2018, revised 5 September 2018, accepted 24 September 2018
Associate Editor: Alan Lee

APPENDIX B: CONSENT FORMS

INFORMED CONSENT FORM

NORTH-WEST UNIVERSITY RESEARCH PROJECT: Lovebird genotyping

BACKGROUND AND AIMS OF THE STUDY

Problem statement:
Exotic bird breeding in South Africa and worldwide is becoming a growing industry. Breeders export or import chicks to or from foreign countries and often these birds are not accompanied by scientifically verified pedigree information. Buyers also need proof that a chick is a heterozygote of a specific recessive colour trait. Often birds are sold on the basis that the parents are of a certain colour so the chick automatically has to be a heterozygote of that colour, given that the trait is inherited as autosomal recessive. However if the pedigree of the chic cannot be verified, it is impossible to guarantee that the chick is a heterozygote. There is currently no species-specific panel of genetic markers available to verify parentage for exotic bird species, nor has the polymorphisms linked to colour variations been identified. In order to ensure accurate documentation accompanying imported or exported birds, it is impeccable that a parentage verification test is available and that the polymorphisms causing colour variations are identified.

Aim:
The aim of the study is to identify Single Nucleotide Polymorphisms (SNPs) markers and construct a panel of markers for Agapornis (lovebird) birds for application of parentage verification.

Objectives:

- Compile a panel of SNP markers best suited for parentage verification
- Verifying power of constructed panel in known family trios

EXPECTED DURATION OF THE PROJECT

This project is expected to be completed by the end of 2018.

PROCEDURES FOLLOWED DURING THIS PROJECT

The project will use only blood samples on blood cards as well as feather samples.

For breeders who wish to send blood cards, a test kit consisting of a blood card and needle and full instructions will be send to the breeder or owner for sampling. The process involves pricking the bird's toe and collecting only a few drops of blood.

The study will be explained to you and you will be allowed time to ask questions and decide whether you would prefer to give consent to participate.

After giving consent, the genetic data of your bird will be obtained from the participating laboratory or a laboratory assigned to the project laboratory. This laboratory will conduct DNA genotyping of selected SNP markers on the samples and the data will be analysed with the aim and objectives of the study in mind.

Selected birds should preferably be from family structures. It is important to the researchers to have the full pedigree of the birds selected for this study. Please include this information with the consent form.

After all the data have been analysed, the findings will be published in scientific journals, PhD thesis, as well as popular journals.

The knowledge gained from this project will be used by North-West University, Potchefstroom, South Africa, to develop a test to determine bird parentage.

You will receive a report on your bird samples once the study is complete.

DECLARATION OF CONFIDENTIALITY

All the Information provided for the purpose of this study will be treated as highly confidential. Only individuals of the research group and officials, who are entitled to it by law, will have access to information. Data published in a scientific journal will include no information that could identify you or your bird any way.

I understand that I am entitled to withdraw from the project at any time without penalty or loss of benefits. My participation in this project is therefore on a voluntary basis until I request withdrawal in writing.

THE ABOVE-MENTIONED PROJECT WAS THOROUGHLY EXPLAINED AND THE FOLLOWING ADDITIONAL INFORMATION POINTED OUT TO ME

- That the data is used for research purposes and no compensation for this will be provided.
- That I will be kept informed of significant developments in the project.
- That the withdrawal clause was explained to me and that I understand its implications.

ENQUIRIES ABOUT THIS PROJECT SHOULD BE DIRECTED TO:

- General enquiries should be directed to: Dr Rencia van der Sluis, Centre for Human Metabonomics, North-West University, Potchefstroom, South Africa, 2531. Tel +27 18 2992318, E-mail: Rencia.vanderSluis@nwu.ac.za

That I will receive a copy of this informed consent form.

DECLARATION OF CONSENT

I,

_____
(Print full name and surname)

_____
(Name of breeder/enterprise)

_____
(Contact details)

_____

_____
(Bird identities to be used for the study, please also attach pedigree data)

Hereby consent to the participation in the above-mentioned project.

I confirm that I am fully aware of the content of this form and that by signing it I give the necessary consent.

Signed at: _____ on _____.
(Place) (Date)

_____
(Signature of responsible who explained the project and informed consent form to the participant)

APPENDIX C: CONFERENCE PROCEEDINGS, ABSTRACTS AND POSTERS PRESENTED.

The work performed during this project was presented at three conferences and the abstracts, posters and conference proceedings published there is given here.

1. 36TH INTERNATIONAL SOCIETY OF ANIMAL GENETICS (ISAG) CONFERENCE IN DUBLIN, IRELAND. 16 –21 JULY 2017

ORAL PRESENTATIONS

Animal Forensic Genetics Workshop

1 *De novo* genome assembly of *Agapornis roseicollis* and SNP discovery for parentage verification. H. van der Zwan*[1], R. van der Sluis[1], and C. Visser[2], [1]*North-West University, Potchefstroom, North-West, South Africa;* [2]*University of Pretoria, Pretoria, Gauteng, South Africa.*

The African parakeet *Agapornis*, or lovebirds, are globally bred as pets. The main breeding selection criterion is plumage coloration. Birds with rare coloration and their heterozygous offspring (with wildtype coloration) are sold at a premium. Currently, there is no genetic test inferring parentage of the heterozygous offspring, nor has the genes or mutations linked to colour variation been identified. The aim of this study was to discover SNPs to develop a SNP-based parentage verification test for *A. roseicollis* by sequencing, assembling and annotating its *de novo* genome. One young male was selected and its genome sequenced at 100x coverage on the Illumina HiSeq platform. The size of the genome was 1.1Gbp and 15 045 genes were identified. This is comparable to other bird species such as the budgerigar (*Melopsittacus undulates)*. The genomes of the male's parents were sequenced at 30x coverage on the same platform. The parents' reads were mapped to the offspring's reference genome using the Burrow-Wheeler aligner. Making use of the command line Linux-based program Genome Analysis Toolkit (GATK), variants were discovered using the HaplotypeCaller module. Approximately 1.2 million raw SNPs were discovered for the mother and 800 000 for the father. Since it is a non-model organism, hard filtering parameters had to be applied to first extract SNPs and then indels from the raw SNP set. SNPs discovered in indels were discarded. These SNPs were further filtered to include only variants where both parents were heterozygous at that locus indicating heterozygosity from all grandparents. This resulted in a total of 614 700 SNPs after all the filters were applied. SNPs not complying with Mendelian inheritance patterns between the parents and offspring were rejected. Where SNPs was located less than 100bp apart only the SNP with the highest quality score was included. A panel of the top 400 SNPs was selected based on the quality scores obtained during the GATK analyses. This panel will form the foundation for the development of a commercial parentage verification test for lovebirds.

Key Words: lovebird, avian genomics, bioinformatics, *de novo* genome assembly, genomic selection

3 Effectiveness of SNPs genotyping assay as a tool for genetic traceability of cattle production chain. A. Pozzi*[1], C. Previtali[1], R. Capoferri[1], S. Arabi[1], A. Galli[2], and G. Bongioni[1], [1]*Istituto Sperimentale Italiano L. Spallanzani, Rivolta d'Adda, Cremona, Italy;* [2]*Centro di ricerca per le produzioni foraggere e lattiero-casearie CREA, Lodi, Italy.*

In the last years the concern of consumer about food safety is increased, so in the European Union great value is placed on accurate and safe animal identification; breeding plan, disease monitoring, and food safety depend heavily on conventional animal identification. For this reason, the protection of the integrity of these programs from fraud or accidental errors is necessary. Genetic traceability is presented here as a way to further enhance conventional traceability; it represents a powerful tool to verify the identity of animal products during every step of the cattle production chain. From the molecular point of view, Single Nucleotide Polymorphism (SNP) markers have gradually replaced microsatellites (STRs), mostly due to their abundance, cost efficiency, and potential for automation. In this work, the amount of information provided by a 32 SNP panel, selected in a previous study was evaluated on different biological matrices. The main sources of genomic DNA are peripheral blood leukocytes, semen, hair follicles and tissues but sampling requires specialised staff and the animals may be subject to stress, influencing negatively welfare, health and productivity. A valid and cheaper alternative of genomic source can be the milk somatic cells (from 2×10^6 to 2×10^5 cells per millilitre of milk). In details, DNA from 50 semen, 30 meat, 30 hair and 30 milk samples were collected. DNA from 140 samples was processed by TaqMan PCR and scanning array on the Open Array platform, the genotypes were generated by SNP Genotyping and TaqMan Genotyper software. Allele frequencies for each SNP were determined by direct counting. A software developed 'ad hoc' (PAF), was used to estimate the levels of genetic variability: expected heterozygosity (He) values ranging from 0.47 to 0.50 with medium value of 0.46, observed heterozygosity (Ho) ranging from 0,38 to 0,60 and medium value of 0,48. Probability of identity (PI) was calculated for the 32 SNPs and it was equal to 8.40×10^{-14} The 32 SNPs assay described in this study represents a valid and useful tool for DNA-based traceability employed in different applied research projects and in the major commercial cattle products.

Key Words: cattle, single nucleotide polymorphisms, product traceability

4 The population and landscape genetics of the European badger (*Meles meles)* in Ireland. A. Allen*[1], J. Guerrero[2], A. Byrne[1,3], J. Lavery[1], E. Presho[1], G. Kelly[1], E. Courcier[4], J. O'Keefe[5], U. Fogarty[6], D. O'Meara[7], G. Wilson[8], D. Ensing[1], C. McCormick[1], R. Biek[9], R. Skuce[1,3], [1]*Agri Food and Biosciences Institute, Belfast, Northern Ireland;* [2]*CEFE-CNRS, Centre D'Ecologie Fonctionelle et Evolutie, Montpelier, France;* [3]*School of Biological Sciences, Queen's University Belfast, Belfast, Northern Ireland;* [4]*Department of Agriculture, Environment and Rural Affairs, Belfast, Northern Ireland;* [5]*Department of Agriculture Food and the Marine, Dublin, Ireland;* [6]*Irish Equine Centre, Johnstown, Republic of Ireland;* [7]*Waterford Institute of Technology, Waterford, Ireland;* [8]*Animal and Plant Health Agency (APHA), Stonehouse, Gloucestershire, England;* [9]*University of Glasgow, Glasgow, Scotland.*

The European badger (*Meles meles*) is an important member of the fauna of Britain and Ireland, not least because it acts as a wildlife reservoir for bovine tuberculosis. Genetic structure of the species is expected to have been influenced by anthropogenic activities and also landscape-level effects. The relative contribution of both factors is debated, but will conceivably have implications for both wildlife and disease management. Recent Europe-wide surveys of genetic diversity have suggested human-aided introduction of badgers into Ireland. These studies have not, however, indexed island-wide diversity of the species, nor comprehensively attempted to detail demographic and geographic factors which shaped the extant population. Herein, we detail the most comprehensive population and landscape genetic study of the badger in Ireland to date. Our data demonstrate that north-eastern and south-eastern counties of Ireland contain a badger sub-population genetically similar to its British contemporaries. Approximate Bayesian computation suggests this sub-population arose in Ireland ~250–3500 years ago through likely import of a small number of badgers from Britain, which then admixed with an already resident Irish badger population. Landscape genetic analyses determined that geographic distance and elevation were the primary drivers of genetic differentia-

POSTER PRESENTED:

De novo genome assembly of *Agapornis roseicollis* and SNP discovery for parentage verification

H. Van der Zwan *[1], C. Visser[2], R. van der Sluis[1]
[1]North-West University, Potchefstroom, South Africa [2]University of Pretoria, Pretoria, South Africa

Agapornis roseicollis breeding

- African parrot with nine species commonly called lovebirds
- Best known as pets but wild flocks still found
- Many colour variations known which is main selection criterion
- No reference genome or genetic screening tests available

Objectives of the study

- Sequence the *de novo* genome of *A. roseicollis*
- Assemble and annotate the genome
- Sequence the parents of reference bird at a lower coverage
- Utilise GATK to discover SNPs from parents' genomes
- Filter and refine SNPs

De novo sequencing of genome

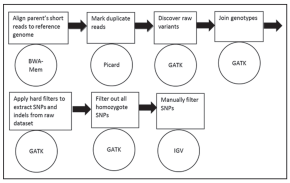
- Young male born and bred in captivity in Belgium selected as reference bird
- Sequence genome at 100x coverage at Eurofins Genomics, Germany
- Sequenced on Illumina HiSeq 2000 platform with chemistry 4

De novo assembly and annotation

- Assembly and annotation done at BGI
- Collaboration with B10K and the Avian Phylo-genomics project
- SOAPdenovo v2.04 used as assembler
- Gene annotation done with RepeatMasker 4.0.5 and Ensembl gene sets
- GenBank accession number: NDXB01000000

SNP Discovery Methods

- Parents of reference bird sequenced at 30x coverage
- Parents' short reads mapped to reference genome of offspring to discover SNPs to be included in parentage verification panel



Results

Comparative Avian Genome Results:

Organism	Genome size (bp)	Scaffold N50	Contig N50	Number of genes	Sequence depth	Reference
Agapornis roseicollis (Lovebird)	1.1G	109K	5.4K	16 044	100x, Illumina	Current study
Melopsittacus undulatus (Budgerigar)	1.2G	10.6M	55.6K	16 240	160x, Multiple platforms	Ganapathy *et al.* (2014)
Gallus gallus (Chicken)	1.05G	11.1M	45.28K	17 108	7x, Sanger	Hillier *et al.* (2004)
Taeniopygia guttata (Zebra finch)	1.2G	10.4M	38.5K	18 618	6x, Sanger	Warren *et al.* (2010)
Amazona vittata (Puerto Rican parrot)	1.58G	19.47K	6.9K	N/A	27x, Illumina	Oleksyk *et al.* (2012)
Ara macao (Scarlet macaw)	1.2G	15.9K	6.3K	14 405	16x, Roche 454; 13x, Illumina	Seabury *et al.* (2013)

SNP discovery results:

	Mother	Father	Combined
Raw variants	2 156 950 variants	1 601 584 variants	N/A
SNPs only	1 916 289 SNPs	1 428 869 SNPs	1 667 639 SNPs
Heterozygote SNPs	N/A	N/A	1 060 242 SNPs

Conclusion and Future Projects

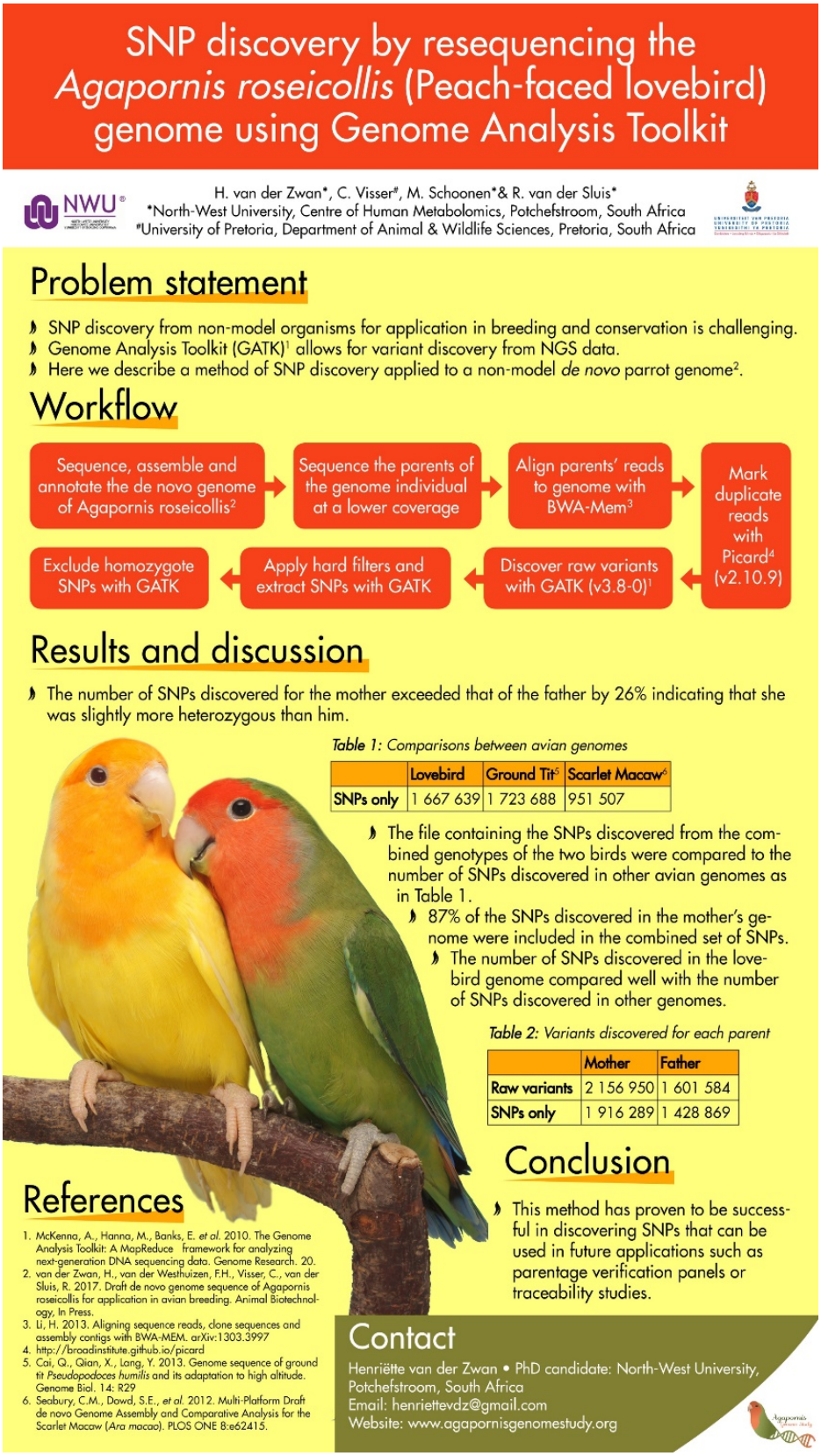
- Genome of *A. roseicollis* was sequenced, assembled and annotated at a sufficient quality to utilize for future gene function projects
- SNPs discovered during the project will be filtered further and reduced to a panel of 400 SNPs
- This panel will be tested in a larger population to finalise a parentage and individual verification panel for lovebirds
- Identification of causative variants associated with psittacofulvin pigment production

Contact

Henriëtte van der Zwan
PhD candidate: North-West University, Potchefstroom, South Africa
Email: henriettevdz@gmail.com
Website: www.agapornisgenomestudy.org

2. 11TH WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION (WCGALP). AUCKLAND, NEW ZEALAND. 11 – 16 FEBRUARY 2018.

POSTER PRESENTED:



SNP discovery by resequencing the *Agapornis roseicollis* (Peach-faced lovebird) genome using Genome Analysis Toolkit

H. van der Zwan*, C. Visser#, M. Schoonen*& R. van der Sluis*
*North-West University, Centre of Human Metabolomics, Potchefstroom, South Africa
#University of Pretoria, Department of Animal & Wildlife Sciences, Pretoria, South Africa

Problem statement

- SNP discovery from non-model organisms for application in breeding and conservation is challenging.
- Genome Analysis Toolkit (GATK)[1] allows for variant discovery from NGS data.
- Here we describe a method of SNP discovery applied to a non-model *de novo* parrot genome[2].

Workflow

Sequence, assemble and annotate the de novo genome of Agapornis roseicollis[2] → Sequence the parents of the genome individual at a lower coverage → Align parents' reads to genome with BWA-Mem[3] → Mark duplicate reads with Picard[4] (v2.10.9) → Discover raw variants with GATK (v3.8-0)[1] → Apply hard filters and extract SNPs with GATK → Exclude homozygote SNPs with GATK

Results and discussion

- The number of SNPs discovered for the mother exceeded that of the father by 26% indicating that she was slightly more heterozygous than him.

Table 1: Comparisons between avian genomes

	Lovebird	Ground Tit[5]	Scarlet Macaw[6]
SNPs only	1 667 639	1 723 688	951 507

- The file containing the SNPs discovered from the combined genotypes of the two birds were compared to the number of SNPs discovered in other avian genomes as in Table 1.
  - 87% of the SNPs discovered in the mother's genome were included in the combined set of SNPs.
  - The number of SNPs discovered in the lovebird genome compared well with the number of SNPs discovered in other genomes.

Table 2: Variants discovered for each parent

	Mother	Father
Raw variants	2 156 950	1 601 584
SNPs only	1 916 289	1 428 869

Conclusion

- This method has proven to be successful in discovering SNPs that can be used in future applications such as parentage verification panels or traceability studies.

References

1. McKenna, A., Hanna, M., Banks, E. *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20.
2. van der Zwan, H., van der Westhuizen, F.H., Visser, C., van der Sluis, R. 2017. Draft de novo genome sequence of Agapornis roseicollis for application in avian breeding. Animal Biotechnology, In Press.
3. Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997
4. http://broadinstitute.github.io/picard
5. Cai, Q., Qian, X., Lang, Y. 2013. Genome sequence of ground tit *Pseudopodoces humilis* and its adaptation to high altitude. Genome Biol. 14: R29
6. Seabury, C.M., Dowd, S.E., *et al.* 2012. Multi-Platform Draft de novo Genome Assembly and Comparative Analysis for the Scarlet Macaw (*Ara macao*). PLOS ONE 8:e62415.

Contact

Henriëtte van der Zwan • PhD candidate: North-West University, Potchefstroom, South Africa
Email: henriettevdz@gmail.com
Website: www.agapornisgenomestudy.org

3. SASBMB – SASBMB CONFERENCE, POTCHEFSTROOM, SOUTH AFRICA. 8 -11 JULY 2018.

The abstract submitted was selected for an oral presentation.

ABSTRACT SUBMITTED: THE APPLICATION OF BIOTECHNOLOGY ON GENUS AGAPORNIS (LOVEBIRDS).

Henriette Van der Zwan[1], Rencia van der Sluis[1]; Carina Visser[2]
North-West University, South Africa[1]
University of Pretoria, South Africa[2]

The genus Agapornis consists of nine small African parrot species that are globally well known as pets, but are also found in their native habitat. Illegal trappings, poaching and habitat destruction are the main threats these birds face in the wild. In aviculture, Agapornis breeding is highly popular all across the globe. Birds are mainly selected based on their plumage colour variations but very little molecular research has been conducted on this topic. There are 30 known colour variations amongst the nine species and most of these are inherited as Mendelian traits. However, to date none of the genes or polymorphisms linked to these variations have been identified or verified. Due to unethical breeding practices the need for the development of molecular tests such as identification verification tests or speciation tests is growing. Future research is paramount to ensure the conservation of wild populations as well as aiding breeders in improving breeding strategies.

APPENDIX D: MATERIALS AND METHODS AS WELL AS RESULTS NOT SHOWN IN PAPER II.

A young *A. roseicollis* male of one year of age was selected to be the reference genome individual. Figure D.1. shows the blood collection process of this bird. DNA sequencing, assembly and annotation of the genome are described in Published Paper II, but additional materials, methods and results not covered in the paper is shown here.

Sequencing results generated during the sequencing of the reference genome individual are given in Table D.1.

Table D.1: Sequencing statistics of the reference genome individual sequenced at Eurofins Genomics.

Library	Yield (Mbp)[#]	%Q30[§] (%)	MeanQ
300bp_lane1	33 970	90.98	34.70
300bp_lane2	33 319	90.92	34.68
550bp_lane1	14 955	86.79	33.61
550bp_lane2	14 723	86.64	33.57
750bp_lane1	11 824	84.63	33.14
750bp_lane2	11 717	84.44	33.07
3kb_LJD	29 934	87.45	33.48
8kb_LJD	13 426	86.73	33.29

#: number of bases called in mega bases.
§: represents the percentage of bases with a quality score of at least 30 (inferred base call accuracy of 99.9%). The Q-score is a prediction of the probability of a wrong base call.

Read lengths of the long jumping distance (LJD) pairs generated in the lovebird study had a mean length of 95bp whereas unknown pairs had a mean length of 122 bp, paired end pairs 102 bp and singleton reads a length of 98 bp.

Figure D.1: Blood sampling of the reference genome bird.
Whole blood in an EDTA tube was taken by a veterinarian after the breeder's consent.
Photo: Henriëtte van der Zwan.

The genome assembly was performed at BGI in Copenhagen as part of the avian B10K phylogenomics project. The assembly is described in Published Paper II, but more detail on the methods and some results not shown in the paper are presented here. The lovebird genome was assembled using SOAPdenovo 2.04 that make use of the "overlap-layout-consensus" method and is executed in six modules that construct and assemble both contigs and scaffolds; correct read errors; construct de Bruijn graphs; map paired-end reads and close gaps (Lou *et al.*, 2012). The first step was to construct de Bruijn graphs by splitting corrected-reads with short insert-sizes into *k*-mers. *K*-mers were merged by clipping tips, merging bubbles and removing low coverage links. Various *k*-mer lengths were assessed to infer the contig and scaffold N50 lengths (Table D.2) before a final decision of a *k*-mer length of 69 was made. The final *k*-mer length of 69 was made based on iterations of different lengths in order to find the longest N50 metrics. In Table D.2 different *k*-mer lengths are shown with the resulting N50 lengths and total size of the genome.

Table D.2: Different k-mer lengths and the resulting N50 statistics that were used.

K-mer length	Contig N50 (bp)	Scaffold N50 (bp)	Total size (bp)	Mapped ratio
23	2,158	57,861	1,194,678,165	88.70%
29	3,471	66,319	1,165,249,232	91.10%
39	4,512	81,064	1,158,620,191	92.30%
41	4,684	84,873	1,158,667,495	92.40%
47	5,174	94,613	1,159,816,267	92.50%
51	5,332	100,492	1,177,395,268	92.60%
57	5,455	108,514	1,200,995,334	92.40%

All reads were mapped to contig sequences to construct scaffolds starting from small libraries (300 bp) to the larger ones (550 and 700 bp) using the paired-end information. At least three read pairs were required to form a connection. Gaps between scaffolds were closed by using paired-end information by aligning one end of a paired read to a contig and the other end within a gap. Gaps within scaffolds were filled by the Gapcloser module in SOAPdenovo.

A 17-mer analysis was performed to assess the complexity of the genome and for this analysis the 300 bp insert library was used. The distribution of the 17-mer frequency was normal and didn't indicate any abnormally high heterozygosity and no high repetitive content. Genome size was predicted during this analysis and was estimated at 1.19 Gbp. Figure D.2 shows a graph that depicts the 17-mer analysis.

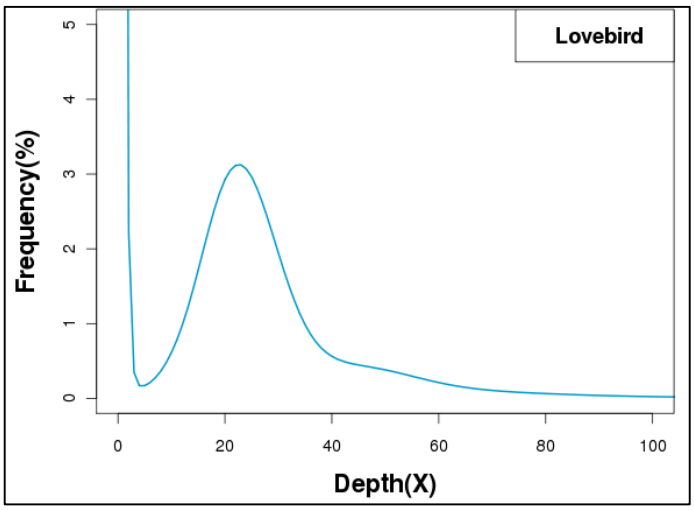


Figure D.2: 17-mer analysis to assess the complexity of the genome.

After assembling the genome, protein coding sequences, repeat sequences and other elements in the genome needed to be identified and named in a process called genome annotation (Stein, 2001; Yandell & Ence, 2012). Annotation was described in much detail in Published Paper II. In Table D.3 the annotation statistics of the Lovebird *de novo* genome annotation that was not shown in the published paper, is given.

Table D.3: Annotation statistics of the Lovebird genome

Gene Set	Gene number	Average mRNA Length (bp)	Average CDS Length (bp)	Exon number	Average Exon Length (bp)	Average Intron Length (bp)
Gallus gallus	16516	348263436	21086	1433	149688	158177
Taeniopygia guttata	17471	373133116	21357	1383	157442	153172
Lovebird	15045	219555241	14593	1389	121862	171195

APPENDIX E: SUPPLEMENTARY FILES SUBMITTED AS PART OF SUBMITTED MANUSCRIPT I (DEVELOPMENT OF A SNP-BASED PARENTAGE VERIFICATION PANEL FOR LOVEBIRDS).

SUPPLEMENTARY FILE 1: COMMAND LINE ARGUMENTS USED DURING GATK PIPELINE.

BWA: Align reads

bwa mem -M -t 16 ref.fa read1.fq read2.fq > aln.sam

Picard: Mark duplicates

java -jar picard.jar MarkDuplicatesWithMateCigar \

I=input.bam \

O=mark_dups_w_mate_cig.bam \

M=mark_dups_w_mate_cig_metrics.txt

GATK: Call variants

java -jar GenomeAnalysisTK.jar \

-T HaplotypeCaller \

-R reference.fa \

-I preprocessed_reads.bam \

--genotyping_mode DISCOVERY \

-stand_emit_conf 10 \

-stand_call_conf 30 \

-o raw_variants.vcf

GATK: Join genotypes

java –jar GenomeAnalysisTK.jar

–T GenotypeGVCFs \

–R abc.fasta \

–V sample1.g.vcf \

–V sample2.g.vcf \

–V sampleN.g.vcf \

–o output.vcf

GATK: Filter variants

Extract the SNP from the call set

```
java -jar GenomeAnalysisTK.jar \
-T SelectVariants \
-R reference.fa \
-V raw_variants.vcf \
-selectType SNP \
-o raw_snps.vcf
```

Extract indels from the call set

```
java -jar GenomeAnalysisTK.jar \
-T SelectVariants \
-R reference.fa \
-V raw_HC_variants.vcf \
-selectType INDEL \
-o raw_indels.vcf
```

SUPPLEMENTARY FILE 2: VERIFICATION OF NGS SNP DATA USING SANGER SEQUENCING.

	SNP_500		SNP_600		SNP_700		SNP_549		SNP_650	
	NGS*	SS#	NGS	SS	NGS	SS	NGS	SS	NGS	SS
Father	AG	AG	CT	CT	AC	AC	AG	AG	AG	AG
Chick	G	AG	C	CC	A	AA	G	GG	G	GG
A. fisheri	N/A	AA	N/A	CC	N/A	AA	N/A	GG	N/A	GG
A. taranta	N/A	GG	N/A	CC	N/A	AA	N/A	GG	N/A	GG
	SNP_801		SNP_850		SNP_903		SNP_951		SNP_752	
	NGS	SS	NGS	SS	NGS	SS	NGS	SS	NGS	SS
Father	AT	AT	CT	CT	AC	AC	CT	CC	CC	CC
Chick	A	AA	C	CT	C	CC	C	CC	C	CC
A. fishceri	N/A	TT	N/A	TT	N/A	AA	N/A	N/A	N/A	CC
A. taranta	N/A	TT	N/A	TT	N/A	AA	N/A	N/A	N/A	CC

*NGS: Next Generation Sequencing Data

#SS : Sanger Sequencing Data

SUPPLEMENTARY FILE 3: MINOR ALLELE FREQUENCIES (MAF) AND OBSERVED HETEROZYGOSITY (H_O) VALUES OF THE SNPS AS COMPILED IN THE THREE PANELS.

262 SNP Panel			195 SNP Panel			40 SNP panel		
Locus	MAF	HObs	Locus	MAF	Hobs	Locus	MAF	HObs
SNP_14	0.5000	1.000	SNP_14	0.5000	1.000	SNP_14	0.5000	1.000
SNP_549	0.5000	1.000	SNP_549	0.5000	1.000	SNP_549	0.5000	1.000
SNP_597	0.5000	1.000	SNP_597	0.5000	1.000	SNP_597	0.5000	1.000
SNP_624	0.5000	1.000	SNP_624	0.5000	1.000	SNP_624	0.5000	1.000
SNP_668	0.5000	1.000	SNP_668	0.5000	1.000	SNP_668	0.5000	1.000
SNP_763	0.5000	1.000	SNP_763	0.5000	1.000	SNP_763	0.5000	1.000
SNP_804	0.5000	1.000	SNP_804	0.5000	1.000	SNP_804	0.5000	1.000
SNP_829	0.5000	1.000	SNP_829	0.5000	1.000	SNP_829	0.5000	1.000
SNP_131	0.4774	0.940	Hom_SNP_11	0.4915	0.139	SNP_131	0.4774	0.940
SNP_713	0.4756	0.896	SNP_652	0.4851	0.888	SNP_713	0.4756	0.896
SNP_652	0.4851	0.888	SNP_814	0.4831	0.151	SNP_652	0.4851	0.888
SNP_683	0.4671	0.860	SNP_131	0.4774	0.940	SNP_683	0.4671	0.860
SNP_785	0.4360	0.849	SNP_713	0.4756	0.896	SNP_785	0.4360	0.849
SNP_827	0.4238	0.848	SNP_124	0.4705	0.806	SNP_827	0.4238	0.848
SNP_108	0.4457	0.808	SNP_683	0.4671	0.860	SNP_108	0.4457	0.808
SNP_124	0.4705	0.806	SNP_129	0.4664	0.114	SNP_124	0.4705	0.806
SNP_517	0.4501	0.753	SNP_761	0.4601	0.526	SNP_517	0.4501	0.753
SNP_107	0.4597	0.752	SNP_107	0.4597	0.752	SNP_107	0.4597	0.752
SNP_566	0.3738	0.748	SNP_137	0.4549	0.141	SNP_566	0.3738	0.748
SNP_170	0.4425	0.725	SNP_536	0.4544	0.576	SNP_170	0.4425	0.725
SNP_727	0.3578	0.716	SNP_132	0.4520	0.328	SNP_727	0.3578	0.716
SNP_673	0.4108	0.697	SNP_517	0.4501	0.753	SNP_673	0.4108	0.697
SNP_160	0.3657	0.684	SNP_548	0.4463	0.640	SNP_160	0.3657	0.684
SNP_649	0.3456	0.647	SNP_108	0.4457	0.808	SNP_649	0.3456	0.647

SNP_548	0.4463	0.640	SNP_805	0.4453	0.297	SNP_548	0.4463	0.640
SNP_581	0.3516	0.607	SNP_839	0.4444	0.267	SNP_581	0.3516	0.607
SNP_757	0.4410	0.578	SNP_731	0.4443	0.437	SNP_757	0.4410	0.578
SNP_536	0.4544	0.576	SNP_170	0.4425	0.725	SNP_536	0.4544	0.576
SNP_639	0.3239	0.575	SNP_757	0.4410	0.578	SNP_639	0.3239	0.575
SNP_144	0.2857	0.563	SNP_785	0.4360	0.849	SNP_102	0.2990	0.558
SNP_102	0.2990	0.558	SNP_622	0.4345	0.291	SNP_761	0.4601	0.526
SNP_761	0.4601	0.526	SNP_759	0.4319	0.197	SNP_731	0.4443	0.437
SNP_118	0.2804	0.506	SNP_825	0.4263	0.152	SNP_100	0.3352	0.433
SNP_106	0.2618	0.503	SNP_827	0.4238	0.848	SNP_132	0.4520	0.328
SNP_138	0.2799	0.463	SNP_725	0.4120	0.193	SNP_172	0.3967	0.319
SNP_127	0.2609	0.454	SNP_673	0.4108	0.697	SNP_136	0.2935	0.315
SNP_731	0.4443	0.437	SNP_701	0.4023	0.114	SNP_559	0.3677	0.311
SNP_100	0.3352	0.433	SNP_172	0.3967	0.319	Hom_SNP_7	0.3773	0.308
SNP_820	0.2830	0.430	SNP_105	0.3962	0.234	SNP_805	0.4453	0.297
SNP_633	0.2225	0.413	SNP_744	0.3842	0.119	SNP_622	0.4345	0.291
SNP_718	0.2546	0.403	Hom_SNP_7	0.3773	0.308			
SNP_494	0.1938	0.388	SNP_133	0.3754	0.206			
SNP_738	0.1986	0.374	SNP_612	0.3744	0.144			
SNP_734	0.1799	0.331	SNP_566	0.3738	0.748			
SNP_132	0.4520	0.328	SNP_739	0.3679	0.119			
SNP_172	0.3967	0.319	SNP_559	0.3677	0.311			
SNP_136	0.2935	0.315	SNP_160	0.3657	0.684			
SNP_567	0.2053	0.314	SNP_743	0.3620	0.181			
SNP_559	0.3677	0.311	SNP_696	0.3604	0.129			
SNP_598	0.1849	0.309	SNP_727	0.3578	0.716			
SNP_794	0.1821	0.309	SNP_581	0.3516	0.607			
Hom_SNP_7	0.3773	0.308	SNP_573	0.3516	0.105			
SNP_805	0.4453	0.297	SNP_556	0.3503	0.269			

SNP_622	0.4345	0.291	SNP_782	0.3467	0.147			
SNP_724	0.1584	0.287	SNP_649	0.3456	0.647			
SNP_142	0.2924	0.284	SNP_552	0.3394	0.124			
SNP_667	0.2690	0.277	SNP_799	0.3367	0.220			
SNP_556	0.3503	0.269	SNP_611	0.3361	0.188			
SNP_839	0.4444	0.267	SNP_100	0.3352	0.433			
SNP_109	0.1506	0.266	SNP_139	0.3352	0.257			
SNP_638	0.2674	0.265	SNP_831	0.3312	0.131			
SNP_15	0.2245	0.265	SNP_726	0.3290	0.101			
SNP_139	0.3352	0.257	SNP_639	0.3239	0.575			
SNP_524	0.2861	0.257	SNP_632	0.3230	0.139			
SNP_694	0.3016	0.255	SNP_810	0.3227	0.148			
Hom_SNP_12	0.2368	0.242	SNP_693	0.3223	0.116			
SNP_150	0.2669	0.240	SNP_576	0.3166	0.128			
SNP_514	0.1189	0.238	SNP_762	0.3133	0.136			
SNP_105	0.3962	0.234	SNP_521	0.3111	0.157			
SNP_500	0.1231	0.231	SNP_791	0.3030	0.164			
SNP_976	0.1974	0.226	SNP_694	0.3016	0.255			
SNP_547	0.1223	0.223	SNP_102	0.2990	0.558			
Hom_SNP_14	0.2059	0.222	SNP_136	0.2935	0.315			
SNP_799	0.3367	0.220	SNP_719	0.2929	0.215			
SNP_719	0.2929	0.215	SNP_606	0.2926	0.173			
SNP_534	0.1856	0.213	SNP_142	0.2924	0.284			
SNP_711	0.2614	0.212	SNP_822	0.2924	0.129			
SNP_133	0.3754	0.206	SNP_571	0.2883	0.149			
SNP_537	0.1829	0.201	SNP_777	0.2868	0.143			
SNP_631	0.1932	0.200	SNP_776	0.2867	0.142			
SNP_630	0.1564	0.199	SNP_524	0.2861	0.257			
SNP_759	0.4319	0.197	SNP_522	0.2859	0.189			

SNP_162	0.1598	0.197	SNP_144	0.2857	0.563			
SNP_579	0.2740	0.196	SNP_820	0.2830	0.430			
SNP_756	0.2354	0.196	SNP_821	0.2818	0.120			
SNP_725	0.4120	0.193	SNP_118	0.2804	0.506			
SNP_686	0.2584	0.190	SNP_138	0.2799	0.463			
SNP_522	0.2859	0.189	SNP_579	0.2740	0.196			
SNP_644	0.2135	0.189	SNP_605	0.2706	0.182			
Hom_SNP_20	0.1578	0.189	SNP_806	0.2704	0.129			
SNP_611	0.3361	0.188	SNP_704	0.2696	0.157			
SNP_516	0.1554	0.188	SNP_667	0.2690	0.277			
SNP_12	0.1794	0.186	SNP_507	0.2677	0.153			
SNP_605	0.2706	0.182	SNP_638	0.2674	0.265			
SNP_743	0.3620	0.181	SNP_150	0.2669	0.240			
SNP_623	0.2664	0.181	SNP_623	0.2664	0.181			
SNP_601	0.1820	0.180	SNP_752	0.2664	0.154			
SNP_17	0.1667	0.180	SNP_675	0.2639	0.166			
SNP_681	0.2623	0.178	SNP_501	0.2633	0.137			
SNP_606	0.2926	0.173	SNP_793	0.2625	0.134			
SNP_634	0.1671	0.173	SNP_681	0.2623	0.178			
SNP_981	0.1215	0.173	SNP_106	0.2618	0.503			
SNP_572	0.2215	0.170	SNP_711	0.2614	0.212			
SNP_596	0.2547	0.168	SNP_811	0.2612	0.155			
SNP_768	0.2287	0.168	SNP_127	0.2609	0.454			
SNP_675	0.2639	0.166	SNP_686	0.2584	0.190			
SNP_119	0.1169	0.166	SNP_596	0.2547	0.168			
SNP_787	0.2430	0.165	SNP_544	0.2547	0.151			
SNP_791	0.3030	0.164	SNP_718	0.2546	0.403			
SNP_515	0.2528	0.163	SNP_515	0.2528	0.163			
SNP_650	0.1551	0.162	SNP_780	0.2486	0.154			

SNP_156	0.1201	0.161	SNP_790	0.2470	0.124			
SNP_558	0.1443	0.159	SNP_787	0.2430	0.165			
SNP_521	0.3111	0.157	SNP_826	0.2380	0.100			
SNP_704	0.2696	0.157	Hom_SNP_12	0.2368	0.242			
SNP_811	0.2612	0.155	SNP_789	0.2367	0.136			
SNP_717	0.2255	0.155	SNP_756	0.2354	0.196			
SNP_707	0.2000	0.155	SNP_613	0.2336	0.126			
SNP_752	0.2664	0.154	SNP_801	0.2327	0.150			
SNP_780	0.2486	0.154	SNP_768	0.2287	0.168			
SNP_754	0.1742	0.154	SNP_717	0.2255	0.155			
SNP_507	0.2677	0.153	SNP_15	0.2245	0.265			
SNP_825	0.4263	0.152	SNP_633	0.2225	0.413			
SNP_535	0.2193	0.152	SNP_572	0.2215	0.170			
SNP_814	0.4831	0.151	SNP_659	0.2202	0.114			
SNP_544	0.2547	0.151	SNP_535	0.2193	0.152			
SNP_801	0.2327	0.150	SNP_735	0.2161	0.114			
SNP_111	0.1966	0.150	SNP_545	0.2153	0.130			
SNP_571	0.2883	0.149	SNP_644	0.2135	0.189			
SNP_810	0.3227	0.148	SNP_833	0.2131	0.136			
SNP_782	0.3467	0.147	SNP_560	0.2080	0.143			
SNP_615	0.1854	0.146	SNP_747	0.2077	0.142			
SNP_538	0.1321	0.146	SNP_657	0.2069	0.121			
SNP_612	0.3744	0.144	Hom_SNP_14	0.2059	0.222			
SNP_540	0.1296	0.144	SNP_567	0.2053	0.314			
SNP_777	0.2868	0.143	SNP_802	0.2040	0.133			
SNP_560	0.2080	0.143	SNP_589	0.2018	0.133			
SNP_776	0.2867	0.142	SNP_707	0.2000	0.155			
SNP_747	0.2077	0.142	SNP_738	0.1986	0.374			
SNP_137	0.4549	0.141	SNP_976	0.1974	0.226			

SNP_658	0.1558	0.140	SNP_493	0.1971	0.127			
SNP_661	0.1329	0.140	SNP_111	0.1966	0.150			
SNP_583	0.1134	0.140	SNP_494	0.1938	0.388			
Hom_SNP_11	0.4915	0.139	SNP_631	0.1932	0.200			
SNP_632	0.3230	0.139	SNP_577	0.1870	0.135			
SNP_125	0.1160	0.138	SNP_527	0.1867	0.106			
SNP_501	0.2633	0.137	SNP_534	0.1856	0.213			
SNP_621	0.1371	0.137	SNP_615	0.1854	0.146			
SNP_762	0.3133	0.136	SNP_598	0.1849	0.309			
SNP_789	0.2367	0.136	SNP_537	0.1829	0.201			
SNP_833	0.2131	0.136	SNP_794	0.1821	0.309			
SNP_577	0.1870	0.135	SNP_601	0.1820	0.180			
SNP_793	0.2625	0.134	SNP_734	0.1799	0.331			
SNP_532	0.1217	0.134	SNP_12	0.1794	0.186			
SNP_802	0.2040	0.133	SNP_542	0.1746	0.131			
SNP_589	0.2018	0.133	SNP_688	0.1743	0.129			
SNP_831	0.3312	0.131	SNP_754	0.1742	0.154			
SNP_542	0.1746	0.131	SNP_640	0.1695	0.123			
SNP_978	0.0923	0.131	SNP_634	0.1671	0.173			
SNP_545	0.2153	0.130	SNP_17	0.1667	0.180			
SNP_696	0.3604	0.129	SNP_778	0.1641	0.120			
SNP_822	0.2924	0.129	SNP_162	0.1598	0.197			
SNP_806	0.2704	0.129	SNP_715	0.1592	0.125			
SNP_688	0.1743	0.129	SNP_724	0.1584	0.287			
SNP_767	0.1270	0.129	Hom_SNP_20	0.1578	0.189			
SNP_576	0.3166	0.128	SNP_630	0.1564	0.199			
SNP_809	0.1139	0.128	SNP_658	0.1558	0.140			
SNP_493	0.1971	0.127	SNP_516	0.1554	0.188			
SNP_613	0.2336	0.126	SNP_650	0.1551	0.162			

SNP_715	0.1592	0.125	SNP_737	0.1541	0.115			
SNP_552	0.3394	0.124	SNP_109	0.1506	0.266			
SNP_790	0.2470	0.124	SNP_578	0.1489	0.114			
SNP_574	0.1172	0.124	SNP_647	0.1456	0.122			
SNP_640	0.1695	0.123	SNP_558	0.1443	0.159			
SNP_647	0.1456	0.122	SNP_654	0.1407	0.106			
SNP_657	0.2069	0.121	SNP_828	0.1379	0.121			
SNP_828	0.1379	0.121	SNP_621	0.1371	0.137			
SNP_821	0.2818	0.120	SNP_661	0.1329	0.140			
SNP_778	0.1641	0.120	SNP_538	0.1321	0.146			
SNP_744	0.3842	0.119	SNP_625	0.1320	0.107			
SNP_739	0.3679	0.119	SNP_540	0.1296	0.144			
SNP_693	0.3223	0.116	SNP_767	0.1270	0.129			
SNP_737	0.1541	0.115	Hom_SNP_15	0.1235	0.114			
SNP_991	0.0846	0.115	SNP_500	0.1231	0.231			
SNP_129	0.4664	0.114	SNP_547	0.1223	0.223			
SNP_701	0.4023	0.114	SNP_532	0.1217	0.134			
SNP_659	0.2202	0.114	SNP_981	0.1215	0.173			
SNP_735	0.2161	0.114	SNP_156	0.1201	0.161			
SNP_578	0.1489	0.114	SNP_514	0.1189	0.238			
Hom_SNP_15	0.1235	0.114	SNP_574	0.1172	0.124			
SNP_114	0.0662	0.112	SNP_119	0.1169	0.166			
SNP_496	0.1057	0.111	SNP_125	0.1160	0.138			
SNP_625	0.1320	0.107	SNP_809	0.1139	0.128			
SNP_527	0.1867	0.106	SNP_583	0.1134	0.140			
SNP_654	0.1407	0.106	SNP_496	0.1057	0.111			
SNP_573	0.3516	0.105						
SNP_726	0.3290	0.101						
SNP_826	0.2380	0.100						

SNP_568	0.1338	0.099						
SNP_492	0.1314	0.099						
SNP_575	0.0830	0.099						
SNP_600	0.0660	0.099						
SNP_513	0.3452	0.098						
SNP_813	0.1743	0.098						
SNP_792	0.1531	0.096						
SNP_636	0.1307	0.096						
SNP_563	0.1308	0.095						
SNP_103	0.0506	0.095						
SNP_840	0.3949	0.093						
SNP_691	0.1727	0.091						
SNP_748	0.1085	0.090						
SNP_741	0.0986	0.090						
Hom_SNP_25	0.1036	0.088						
SNP_541	0.3568	0.087						
SNP_803	0.1491	0.087						
SNP_546	0.1141	0.087						
SNP_665	0.1065	0.087						
SNP_676	0.2331	0.085						
SNP_832	0.0932	0.084						
SNP_836	0.1401	0.083						
SNP_619	0.3840	0.081						
Hom_SNP_9	0.1364	0.077						
SNP_740	0.1844	0.076						
SNP_569	0.1440	0.075						
Hom_SNP_13	0.1365	0.075						
Hom_SNP_10	0.4856	0.074						
SNP_769	0.1233	0.074						

SNP_614	0.4568	0.073						
SNP_703	0.4399	0.073						
SNP_786	0.1323	0.073						
SNP_655	0.3830	0.071						
SNP_648	0.3136	0.070						
SNP_816	0.3133	0.070						
SNP_616	0.4037	0.069						
SNP_503	0.0667	0.069						
SNP_526	0.0636	0.068						
SNP_742	0.4562	0.063						
SNP_699	0.0514	0.063						
SNP_835	0.1023	0.062						
SNP_653	0.4197	0.060						
SNP_660	0.4132	0.059						
SNP_678	0.0804	0.059						
SNP_643	0.0506	0.059						
SNP_585	0.1045	0.054						
SNP_672	0.0535	0.053						
SNP_700	0.1324	0.052						
SNP_135	0.0354	0.051						
SNP_580	0.2133	0.050						
SNP_760	0.0784	0.049						
Hom_SNP_8	0.0642	0.047						
SNP_797	0.0721	0.046						
SNP_990	0.0685	0.046						
SNP_608	0.0557	0.046						
SNP_529	0.0339	0.045						
SNP_670	0.4841	0.040						
Hom_SNP_6	0.0215	0.040						

SNP_795	0.1314	0.032						
SNP_153	0.0087	0.017						
SNP_116	0.0076	0.012						
SNP_819	0.0067	0.006						
Hom_SNP_5	0.4457	0.005						
SNP_565	0.0145	0.000						

SUPPLEMENTARY FILE 4: PARENTAGE VERIFICATION FOR THE 43 FAMILIES USING THE 262-SNP, 195-SNP AND 40-SNP PANELS.

Only father included	Only mother included	Both parents included	Change in +/- LOD between 262 and 195 panels
Families with high number of mismatches indicating incorrect breeder's records.			
Families where a low number of loci were compared.			

262-SNP Analysis

Offspring ID	Loci typed	P1 NEP	P2 NEP	Mother ID	Loci typed	Pair loci compared	Pair loci mismatch	Pair LOD score	Father ID	Loci typed	Pair loci compared	Pair loci mismatch	Pair LOD score	Pair confidence	Trio loci compared	Trio loci mismatch	Trio LOD score	Trio confidence
BS161	255	1.09E-0006	1.09E-0006		0	0	0	0.0	BS162	258	252	33	-1.37	-	0	0	0.0	
BS163	241	1.54E-0022	1.54E-0022		0	0	0	0.0	BS162	258	238	19	-3.37	-	0	0	0.0	
BS56	258	8.56E-0015	8.56E-0015		0	0	0	0.0	BS54	256	252	17	-2.20	-	0	0	0.0	
BS265	195	9.13E-0007	5.11E-0005	BS225	261	194	34	-1.48	BS58	261	195	45	-1.97	-	195	58	-1.26	-
BS264	242	1.50E-0007	3.34E-0006	BS225	261	241	41	-1.79	BS58	261	241	53	-2.34	-	241	75	-1.69	-
BS78	251	2.57E-0007	8.92E-0008	BS77	258	250	33	-1.40	BS76	258	250	38	-1.48	-	250	53	-8.64	-
BS22	258	2.25E-0022	7.23E-0029	BS21	251	248	7	6.52	BS20	254	251	6	7.24	*	251	16	5.22	*
BS226	102	1.25E-0012	2.18E-0011	BS229	261	102	7	7.80	BS224	261	102	5	1.18	*	102	13	3.08	*
BS66	258	1.50E-0024	3.28E-0030	BS65	219	217	6	6.01	BS64	260	256	5	6.34	*	256	19	2.50	*
F105	96	7.58E-0008	5.75E-0007	F103	150	69	3	2.85	F104	128	71	7	-1.40	+	71	9	-1.35	*
BS263	254	1.87E-0007	1.61E-0006	BS225	261	253	43	-1.92	BS58	261	253	57	-2.53	-	253	81	-1.87	-
F138	93	9.18E-0006	3.68E-0006	F136	127	57	2	7.10	F137	89	30	4	-9.67	+	30	5	-1.29	*
BS225	261	5.516E-0015	1.84E-0025	BS229	261	260	1	7.36	BS224	261	260	1	5.650	*	260	3	5.05	*
BS45	260	2.82E-0017	2.77E-0028	BS43	259	258	1	6.93	BS42	262	260	1	7.42	*	260	3	7.32	*
BS16	249	1.07E-0007	1.46E-0008	BS15	253	248	1	4.88	BS17	249	245	1	5.02	*	245	2	5.07	*
BS14	249	6.48E-0007	8.20E-0009	BS16	249	245	1	4.73	BS17	249	245	0	5.28	*	245	7	3.31	*
BS224	261	1.08E-0014	8.41E-0033	BS227	256	255	14	-2.43	BS230	125	125	12	-3.83	-	125	30	-6.85	-
BS229	261	3.20E-0020	2.43E-0029	BS227	256	255	2	6.28	BS230	125	125	19	-5.90	-	125	24	-7.17	-

BS266	259	1.59E-0008	1.55E-0006	BS225	261	258	46	-2.03	BS58	261	258	60	-2.63	-	258	83	-1.85	-
BS72	260	2.39E-0008	2.39E-0008		0	0	0	0.0	BS71	254	253	6	1.74	*	0	0	0.0	
BS28	258	6.08E-0026	6.08E-0026		0	0	0	0.0	BS24	259	257	5	7.98	*	0	0	0.0	
F145	135	1.75E-0007	1.751E-0007	0	0	0	0.0	F144	117	83	5	2.33	2.33	0	0	0.0		
BS31	261	3.91E-0021	3.91E-0021		0	0	0	0.0	BS29	260	259	2	8.25	*	0	0	0.0	
BS71	254	1.98E-0007	1.98E-0007		0	0	0	0.0	BS10	251	246	2	4.61	*	0	0	0.0	
REF	260	7.97E-0018	7.97E-0018		0	0	0	0.0	Father	262	260	1	8.67	*	0	0	0.0	
BS32	257	5.69E-0027	5.69E-0027		0	0	0	0.0	BS29	260	255	1	9.62	*	0	0	0.0	
BS74	247	2.35E-0007	2.35E-0007		0	0	0	0.0	BS73	259	245	1	4.64	*	0	0	0.0	
BS75	255	1.12E-0007	1.12E-0007		0	0	0	0.0	BS73	259	254	1	4.91	*	0	0	0.0	
BS102	252	2.22E-0008	6.52E-0010	BS101	252	247	0	5.89	BS100	250	247	0	5.56	*	247	4	4.36	*
BS91	253	1.63E-0007	7.74E-0009	BS90	247	242	0	5.56	BS89	258	252	0	5.66	*	252	3	4.65	*
BS228	260	3.31E-0015	6.79E-0026	BS229	261	259	0	7.31	BS224	261	259	0	6.07	*	259	3	5.00	*
BS88	254	3.45E-0007	7.206E-0009	BS86	256	251	0	5.44	BS85	253	250	0	5.21	*	250	3	4.47	*
BS99	253	2.70E-0007	2.704E-0007	0	0	0	0.0	BS98	225	223	0	4.98	4.98	0	0	0.0		
BS103	252	7.42E-0008	4.71E-0008	BS101	252	248	0	5.91	BS100	250	245	0	5.86	*	245	0	5.73	*
BS44	260	1.44E-0016	7.07E-0028	BS43	259	257	0	7.49	BS42	262	260	0	7.40	*	260	0	7.97	*
BS164	242	3.19E-0008	3.19E-0008		0	0	0	0.0	BS169	249	236	0	5.34	*	0	0	0.0	
BS85	253	5.71E-0007	5.715E-0007	0	0	0	0.0	BS80	253	247	0	5.39	5.39	0	0	0.0		
BS87	253	2.46E-0007	4.83E-0008	BS86	256	248	0	5.47	BS85	253	248	0	5.22	*	248	0	5.46	*
BS259	224	2.189E-0007	1.52E-0007	BS255	229	202	0	5.06	BS257	253	222	0	5.10	*	222	0	5.07	*
BS260	253	2.20E-0007	3.03E-0008	BS255	229	225	0	5.26	BS257	253	251	0	5.69	*	251	0	5.61	*
BS261	248	5.84E-0008	1.19E-0008	BS255	229	222	0	5.27	BS257	253	245	0	5.54	*	245	0	5.55	*
BS120	176	3.013E-0005	3.01E-0005		0	0	0	0.0										

195-SNP Analysis

Offspring ID	Loci typed	P1 NEP	P2 NEP	Mother ID	Loci typed	Pair loci comp	Pair loci mismatch	Pair LOD score	Father ID	Loci typed	Pair loci comp	Pair loci mismatch	Pair LOD score	Pair conf	Trio loci comp	Trio loci mismatch	Trio LOD score	Trio conf
F105	71	4,51E-05	2,18E-05	F103	108	51	1	8,30E+00	F104	97	51	3	-7,25E-01	*	51	5	-3,52E+00	*
REF	194	1,97E-09	1,97E-09		0	0	0	0,00E+00	Father	195	194	0	4,30E+01	*	0	0	0,00E+00	
BS91	187	7,10E-06	3,37E-07	BS90	184	179	0	4,23E+01	BS89	192	186	0	4,29E+01	*	186	3	3,27E+01	*
BS99	187	1,17E-05	1,17E-05		0	0	0	0,00E+00	BS98	166	164	0	3,86E+01	*	0	0	0,00E+00	
BS102	187	9,66E-07	2,84E-08	BS101	187	182	0	4,52E+01	BS100	185	182	0	4,19E+01	*	182	4	2,99E+01	*
BS103	187	3,23E-06	2,05E-06	BS101	187	183	0	4,54E+01	BS100	185	180	0	4,49E+01	*	180	0	4,36E+01	*
BS224	194	1,53E-10	4,85E-22	BS227	190	189	11	-2,01E+01	BS230	94	94	10	-3,03E+01	-	94	23	-5,06E+01	-
BS225	195	3,69E-11	1,93E-19	BS229	194	194	1	5,18E+01	BS224	194	194	1	4,20E+01	*	194	3	3,29E+01	*
BS226	75	3,25E-09	1,10E-09	BS229	194	75	4	1,16E+01	BS224	194	75	4	8,52E+00	*	75	9	8,06E-01	*
BS228	193	2,26E-11	4,32E-19	BS229	194	192	0	5,32E+01	BS224	194	192	0	4,50E+01	*	192	2	3,43E+01	*
BS229	194	4,82E-15	5,28E-22	BS227	190	189	1	4,68E+01	BS230	94	94	14	-4,09E+01	-	94	17	-4,97E+01	-
BS14	184	2,73E-05	8,31E-07	BS16	185	181	0	3,96E+01	BS17	184	180	0	3,98E+01	*	180	5	2,32E+01	*
BS120	137	1,50E-04	1,50E-04		0	0	0	0,00E+00										
BS22	191	6,51E-16	5,13E-21	BS21	185	182	5	4,30E+01	BS20	188	185	3	5,68E+01	*	185	11	4,00E+01	*
BS28	192	5,41E-20	5,41E-20		0	0	0	0,00E+00	BS24	192	191	3	6,54E+01	*	0	0	0,00E+00	
BS31	194	1,79E-17	1,79E-17		0	0	0	0,00E+00	BS29	193	192	1	6,68E+01	*	0	0	0,00E+00	
BS32	191	4,64E-21	4,64E-21		0	0	0	0,00E+00	BS29	193	189	1	7,38E+01	*	0	0	0,00E+00	
BS44	194	8,50E-14	1,83E-22	BS43	193	192	0	5,94E+01	BS42	195	194	0	5,98E+01	*	194	0	6,11E+01	*
BS45	193	2,41E-13	7,47E-22	BS43	193	192	0	5,63E+01	BS42	195	193	1	5,59E+01	*	193	2	4,98E+01	*
BS56	192	7,39E-11	7,39E-11		0	0	0	0,00E+00	BS54	189	186	13	-1,89E+01	-	0	0	0,00E+00	
BS66	191	1,76E-20	1,96E-23	BS65	161	159	3	5,69E+01	BS64	193	189	5	4,29E+01	*	189	12	2,01E+01	*
BS16	185	5,36E-06	3,80E-06	BS15	189	185	0	4,11E+01	BS17	184	181	0	4,19E+01	*	181	0	4,18E+01	*
BS161	189	3,00E-05	3,00E-05		0	0	0	0,00E+00	BS162	191	186	25	-1,05E+02	-	0	0	0,00E+00	

BS163	182	9,77E-18	9,77E-18		0	0	0	0,00E+00	BS162	191	179	14	-1,80E+01	-	0	0	0,00E+00	
BS164	180	9,90E-07	9,90E-07		0	0	0	0,00E+00	BS169	186	176	0	4,08E+01	*	0	0	0,00E+00	
BS72	193	8,19E-06	8,19E-06		0	0	0	0,00E+00	BS71	188	187	4	1,39E+01	*	0	0	0,00E+00	
BS71	188	8,64E-06	8,64E-06		0	0	0	0,00E+00	BS10	186	181	2	3,25E+01	*	0	0	0,00E+00	
BS74	185	8,99E-06	8,99E-06		0	0	0	0,00E+00	BS73	192	183	1	3,40E+01	*	0	0	0,00E+00	
BS75	190	4,83E-06	4,83E-06		0	0	0	0,00E+00	BS73	192	189	1	3,62E+01	*	0	0	0,00E+00	
BS78	187	8,87E-06	4,63E-07	BS77	193	187	22	-9,99E+01	BS76	192	186	26	-1,03E+02	-	186	39	-6,99E+01	-
BS85	188	1,97E-05	1,97E-05		0	0	0	0,00E+00	BS80	187	182	0	4,08E+01	*	0	0	0,00E+00	
BS87	188	8,50E-06	1,67E-06	BS86	190	183	0	4,23E+01	BS85	188	183	0	3,91E+01	*	183	0	4,15E+01	*
BS88	188	1,51E-05	3,14E-07	BS86	190	185	0	4,14E+01	BS85	188	185	0	3,91E+01	*	185	3	3,16E+01	*
BS259	168	6,05E-06	4,22E-06	BS255	170	152	0	3,99E+01	BS257	188	167	0	3,93E+01	*	167	0	3,90E+01	*
BS260	188	9,61E-06	1,32E-06	BS255	170	166	0	4,01E+01	BS257	188	186	0	4,33E+01	*	186	0	4,24E+01	*
BS261	184	2,17E-06	4,45E-07	BS255	170	164	0	4,07E+01	BS257	188	181	0	4,22E+01	*	181	0	4,22E+01	*
BS263	189	8,18E-06	4,70E-06	BS225	195	189	28	-1,33E+02	BS58	194	188	40	-1,81E+02	-	188	59	-1,47E+02	-
BS264	178	6,46E-06	9,63E-06	BS225	195	178	26	-1,21E+02	BS58	194	177	36	-1,63E+02	-	177	53	-1,29E+02	-
BS265	146	2,09E-05	1,11E-04	BS225	195	146	22	-1,01E+02	BS58	194	146	32	-1,42E+02	-	146	41	-9,66E+01	-
BS266	194	6,94E-07	4,54E-06	BS225	195	194	31	-1,44E+02	BS58	194	193	43	-1,91E+02	-	193	61	-1,45E+02	-
F145	106	1,16E-05	1,16E-05		0	0	0	0,00E+00	F144	84	62	5	-2,50E+00	*	0	0	0,00E+00	
F138	74	1,54E-04	4,24E-05	F136	96	44	1	6,73E+00	F137	68	25	3	-6,24E+00	*	25	4	-9,47E+00	*

40-SNP Analysis

Offspring ID	Loci typed	P1 NEP	P2 NEP	Mother ID	Loci typed	Pair loci compared	Pair loci mismatch	Pair LOD score	Father ID	Loci typed	Pair loci comp	Pair loci mismatch	Pair LOD score	Pair conf	Trio loci comp	Trio loci mismatch	Trio LOD score	Trio conf
F105	17	1,05E-01	7,64E-02	F103	21	12	0	2,65E+00	F104	20	13	0	2,98E+00	*	13	0	3,01E+00	*
REF	40	1,98E-01	1,98E-01		0	0	0	0,00E+00	Father	40	40	0	2,49E-01	+	0	0	0,00E+00	
BS91	36	3,02E-01	1,12E-01	BS90	37	35	0	3,68E+00	BS89	38	35	0	3,72E+00	*	35	1	-2,58E-01	+
BS99	35	2,86E-01	2,86E-01		0	0	0	0,00E+00	BS98	30	29	0	3,19E+00	*	0	0	0,00E+00	
BS102	38	1,77E-01	1,77E-01	BS101	35	34	0	3,17E+00	BS100	36	35	0	3,86E+00	*	35	0	3,21E+00	*
BS103	37	2,08E-01	2,08E-01	BS101	35	34	0	3,06E+00	BS100	36	34	0	4,53E+00	*	34	0	3,77E+00	*
BS224	39	6,37E-02	8,14E-03	BS227	39	38	1	-1,70E+00	BS230	20	20	2	-4,90E+00	-	20	2	-2,17E+00	+
BS225	40	8,40E-02	9,79E-03	BS229	40	40	0	4,43E+00	BS224	39	39	0	4,18E+00	*	39	0	5,25E+00	*
BS226	20	2,82E-01	7,69E-02	BS229	40	20	0	2,71E+00	BS224	39	20	0	1,74E+00	+	20	1	-5,81E-01	+
BS228	38	7,28E-02	4,92E-03	BS229	40	38	0	3,36E+00	BS224	39	37	0	4,59E+00	*	37	0	6,23E+00	*
BS229	40	3,06E-02	3,06E-02	BS227	39	39	0	6,20E+00	BS230	20	20	0	4,08E+00	*	20	0	3,29E+00	*
BS14	34	5,68E-01	1,25E-01	BS16	34	34	0	1,49E+00	BS17	35	34	0	1,80E+00	+	34	3	-8,98E+00	-
BS120	27	3,88E-01	3,88E-01		0	0	0	0,00E+00										
BS22	37	5,77E-02	1,89E-02	BS21	38	36	0	6,89E+00	BS20	36	34	0	7,64E+00	*	34	0	7,22E+00	*
BS28	38	2,29E-02	2,29E-02		0	0	0	0,00E+00	BS24	39	38	1	3,54E+00	*	0	0	0,00E+00	
BS31	39	4,73E-02	4,73E-02		0	0	0	0,00E+00	BS29	39	38	1	3,11E+00	*	0	0	0,00E+00	
BS32	38	3,48E-02	3,48E-02		0	0	0	0,00E+00	BS29	39	37	0	7,19E+00	*	0	0	0,00E+00	
BS44	40	9,90E-02	2,71E-02	BS43	39	39	0	3,96E+00	BS42	40	40	0	4,33E+00	*	40	0	4,89E+00	*
BS45	39	1,13E-01	3,08E-02	BS43	39	39	0	3,87E+00	BS42	40	39	0	4,86E+00	*	39	0	5,08E+00	*
BS56	39	6,86E-02	6,86E-02		0	0	0	0,00E+00	BS54	35	34	2	-4,59E+00	-	0	0	0,00E+00	
BS66	38	3,64E-02	1,56E-02	BS65	33	33	1	-4,61E-01	BS64	39	37	0	4,40E+00	*	37	1	3,92E+00	*
BS16	34	2,33E-01	1,65E-01	BS15	35	34	0	1,72E+00	BS17	35	34	0	2,65E+00	*	34	0	2,54E+00	*
BS161	36	2,60E-01	2,60E-01		0	0	0	0,00E+00	BS162	36	33	1	-3,25E+00	-	0	0	0,00E+00	
BS163	37	5,35E-02	5,35E-02		0	0	0	0,00E+00	BS162	36	34	1	-1,79E-01	+	0	0	0,00E+00	

BS164	34	1,02E-01	1,02E-01		0	0	0	0,00E+00	BS169	36	34	0	3,67E+00	*	0	0	0,00E+00	
BS72	39	5,58E-01	5,58E-01		0	0	0	0,00E+00	BS71	36	36	0	5,14E-01	+	0	0	0,00E+00	
BS71	36	1,90E-01	1,90E-01		0	0	0	0,00E+00	BS10	33	31	1	-1,26E+00	+	0	0	0,00E+00	
BS74	34	3,18E-01	3,18E-01		0	0	0	0,00E+00	BS73	38	33	0	1,11E+00	+	0	0	0,00E+00	
BS75	38	1,70E-01	1,70E-01		0	0	0	0,00E+00	BS73	38	37	0	2,17E+00	+	0	0	0,00E+00	
BS78	34	3,39E-01	6,65E-03	BS77	39	34	0	-1,92E-01	BS76	38	33	0	1,84E+00	+	33	2	-3,37E+00	+
BS85	36	5,99E-01	5,99E-01		0	0	0	0,00E+00	BS80	36	33	0	1,99E+00	+	0	0	0,00E+00	
BS87	36	1,36E-01	5,51E-02	BS86	38	34	0	2,93E+00	BS85	36	33	0	1,73E+00	+	33	0	1,70E+00	*
BS88	35	2,78E-01	7,83E-02	BS86	38	35	0	2,30E+00	BS85	36	34	0	1,89E+00	+	34	0	1,96E+00	*
BS259	31	2,22E-01	1,55E-01	BS255	34	27	0	3,99E+00	BS257	36	30	0	1,34E+00	+	30	0	8,96E-01	*
BS260	36	4,23E-01	6,01E-02	BS255	34	31	0	1,88E+00	BS257	36	35	0	2,50E+00	*	35	0	2,54E+00	*
BS261	37	1,91E-01	3,93E-02	BS255	34	31	0	2,63E+00	BS257	36	35	0	2,46E+00	*	35	0	2,41E+00	*
BS263	35	3,39E-01	2,68E-02	BS225	40	35	1	-5,85E+00	BS58	39	34	2	-9,68E+00	-	34	6	-1,89E+01	-
BS264	32	3,01E-01	4,39E-02	BS225	40	32	1	-5,18E+00	BS58	39	31	1	-4,39E+00	-	31	4	-1,08E+01	-
BS265	25	1,55E-01	9,67E-02	BS225	40	25	1	-4,25E+00	BS58	39	25	1	-4,83E+00	-	25	1	-1,65E+00	+
BS266	40	2,10E-01	6,33E-03	BS225	40	40	1	-5,62E+00	BS58	39	39	3	-1,30E+01	-	39	7	-2,25E+01	-
F145	19	2,48E-01	2,48E-01		0	0	0	0,00E+00	F144	18	14	1	-1,08E+00	+	0	0	0,00E+00	
F138	16	5,55E-02	4,45E-02	F136	22	13	0	4,56E+00	F137	16	8	2	-6,11E+00	-	8	2	-6,76E+00	-

Pedigree relationships as per breeders				
Chick	Mother	Father		Species
F105	F103	F104		N
REFERENCE		Father		R
BS91	BS90	BS89		P
BS99	BS98			F
BS102	BS101	BS100		N
BS103	BS101	BS100		N
BS224	BS227	BS230		R
BS225	BS229	BS224		R
BS226	BS229	BS224		R
BS228	BS229	BS224		R
BS229	BS227	BS230		R
BS14	BS16	BS17		F
BS120	B119			L
BS22	BS21	BS20		R
BS28	BS24			R
BS31	BS29			R
BS32	BS29			R
BS44	BS43	BS42		R
BS45	BS43	BS42		R
BS56		BS54		R
BS66	BS65	BS64		R
BS16	BS15	BS17		F
BS14	BS15	BS17		F
BS161		BS162		R
BS163		BS162		R
BS164	BS169			F
BS72		BS71		F
BS71		BS10		F
BS74	BS73			F
BS75	BS73			F
BS78	BS77	BS76		F
BS85	BS80			F
BS87	BS86	BS85		F
BS88	BS86	BS85		F
BS259	BS255	BS257		F
BS260	BS255	BS257		F
BS261	BS255	BS257		F
BS263	BS225	BS58		F
BS264	BS225	BS58		F
BS265	BS225	BS58		F
BS266	BS225	BS58		F

F145		F144		P
F138	F136	F137		P
N: A, nigrigenis				
R: A, roseicollis				
F: A, fischeri				
L: A, lilianae				
P: A, personatus				

APPENDIX F: SHORT COMMUNICATION SUBMITTED TO ANIMAL GENETICS

Development of a SNP-based parentage verification panel for lovebirds.

Journal:	*Animal Genetics*
Manuscript ID	Draft
Manuscript Type:	Short Communication
Date Submitted by the Author:	n/a
Complete List of Authors:	Van der Zwan, Henriette; North-West University, Focus Area for of Human Metabolomics Visser, Carina; University of Pretoria, Animal and Wildlife Studies Schoonen, Maryke; North-West University, Focus Area for Human Metabolomics van der Sluis, Rencia; North-West University, Focus Area for Human Metabolomics
Keywords:	parrot breeding, pedigree confirmation, whole genome resequencing, Agapornis

Development of a SNP-based parentage verification panel for lovebirds

Henriëtte van der Zwan*[1], Carina Visser[2], Maryke Schoonen[1] & Rencia van der Sluis[1]
[1]Focus Area for Human Metabolomics, North-West University, Potchefstroom, South Africa
[2]Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa
*Corresponding author Email: agapornis.genome.study@gmail.com

Abstract

Genus *Agapornis*, or lovebirds, are popular pet parrots worldwide. Currently, breeders are dependent on pedigree records as a selection tool as no molecular parentage verification test is available for any of the nine species. The *A. roseicollis* reference genome was recently assembled followed by the sequencing of the whole genomes of the parents of the reference genome individual at 30x coverage. The parents' reads were mapped against the reference genome to identify single nucleotide polymorphisms (SNPs). Over 1.6 million SNPs, shared between the parents, were discovered using the Genome Analysis Toolkit (GATK) pipeline. SNPs were filtered to a panel of 480 SNPs based on GATK parameters. The panel of 480 SNPs was genotyped in a population of 960 lovebirds across seven species. A panel of 262 SNPs were compiled that included SNPs successfully amplified across all species and verified against the father's reference genotype. The 262-SNP panel was reduced based on the Observed Heterozygosity (H_O) and Minor Allele Frequency (MAF) values per SNP to include the lowest number of SNPs with the highest exclusion power for parentage verification. Two smaller panels consisting of respectively 195 SNPs with MAF and H_O values >0.1; and 40 SNPs with MAF and Ho values >0.3, were constructed. The panels were verified using 43 families from different species with known relationships to evaluate the exclusion power of each panel. The 195 SNP panel with an average exclusion probability of 99.9% and MAF and H_O values >0.1 was proposed as the routine *Agapornis* parentage verification panel.

Keywords: *parrot breeding, pedigree confirmation, whole genome resequencing, Agapornis*

The parrot genus *Agapornis* (or lovebirds) consists of nine African parrot species kept worldwide as companion animals (Dilger, 1960; Van den Abeele, 2016) and are also found in their natural habitat across Africa and Madagascar (Forshaw, 1989). Eight of the species are found in captive breeding populations (*A. roseicollis*, *A. fischeri*, *A. lilianae*, *A. nigrigenis*, *A. personatus*, *A. canus, A. pullarius* and *A. taranta*). Three phylogenetic groups are distinguished including the white eye ring group (*A. fischeri*, *A. lilianae*, *A. nigrigenis* and *A. personatus*), intermediate group (*A. roseicollis*), all of which are popular in aviculture, and the sexually dimorphic group (*A. canus, A. pullarius* and *A. taranta*), which are less popular in aviculture (Eberhard, 1998).

Plumage colour is the main selection criterion for breeders and buyers. Most of the 30 observed colour variations are inherited as autosomal or sex-linked recessive traits but the causative mutations associated with these variations have not yet been identified (Van den Abeele, 2016; van der Zwan *et al.*, 2019). Due to the inheritance

patterns of these traits, breeders make use of pedigree data to predict the heterozygous genotype for a specific colour. Therefore, accurate and complete pedigree records are of utmost importance to select offspring. Despite the dependency on pedigree data, no SNP-based, genus-specific parentage verification test is available for *Agapornis* or for any parrot species. As a result, many fraudulent transactions take place as sellers know there is no molecular test available to verify either the pedigree or the colour genotype of a chick.

The development of a SNP-based parentage verification panel that could be applied across all the species of *Agapornis* would be beneficial. In a previous study, the *de novo* genome of *A. roseicollis* was sequenced, assembled and annotated (van der Zwan *et al.*, 2018) (NCBI accession number NDXB01000000). The aim of this study was, therefore, to identify SNPs throughout the genome to be included in a SNP-based parentage verification panel for the most popular domesticated lovebird species.

The genomes of both parents of the reference genome individual were sequenced (NCBI SRA accession number PRJNA355979) and mapped against the reference genome to identify SNPs that were shared between the parents using the variant caller, Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). A total of 1 667 629 SNPs were discovered that were analysed and subsequently reduced to 480 for inclusion in the parentage verification panel. (Supplementary File 1). Biological samples from 960 lovebirds spanning seven species were genotyped using the Quantstudio 12 K Flex Real-time PCR system OpenArray technology (Thermo Fisher Scientific) to determine the level of polymorphism for each SNP amplification (see Supplementary File 1). Included in this dataset was 43 lovebird families with known pedigrees from five species that were used to verify parentage.

Three SNP-based parentage verification panels were constructed based on heterozygosity (H_O), minor allele frequency (MAF) and number of alleles, as described in Supplementary File 1. These panels consisted of 262 SNPs, 195 SNPs (MAF and H_O>0.1) and 40 SNPs (MAF and H_O>0.3) respectively, (MAF and H_O values are shown in Supplementary File 2). The combined exclusion probabilities (P_E) for the first (P_{E1}), second (P_{E2}) and parent pair (P_{EP}) for each of the three panels as well as the mean Heterozygosity of each panel are given in Table 1.

Table 1: The Mean Expected Heterozygosity values and Exclusion probabilities of the three panels.

	SNP Panel size		
Exclusion probability	262	195	40
First parent (P_{E1})	0,999999990	0,99999980	0,9931058
Second parent (P_{E2})	0.9999999999	0.9999999999	0,99969307
Parent pair (P_{EP})	0.9999999999	0.9999999999	0,99999753
Mean H_E	0.34	0.37	0.48

The 262 and 195-SNP panels had higher P_{E1} and P_{E2} values, compared to the 40-SNP panel (Table 1). P_{EP} values were similar (>99.9%) across all three panels. The exclusion of all non-informative SNPs with H_O and MAF values below 0.1 had a small, but positive effect on the mean heterozygosity of the panel. As expected, the inclusion of only SNPs with values in excess of 0.3 had a large effect on the mean H_E. Mean H_E was also calculated per species as shown in Supplementary File 1.

The robustness of the three panels were tested by verifying the pedigree data of 43 lovebird families, using Cervus 3.0 (Marshall *et al.*, 1998) by applying the 262-SNPs, 195-SNPs and 40-SNP panels. One family (*A. lilianae* family) amplified at too few loci for comparison between the chick and parent, and was discarded. Complete parentage verification results and relationships as received from the breeders, are shown in Supplementary File 3. In Table 2 a summary of the parentage verification results with special focus on differences between the allocations of the three panels, is given.

Table 2: Summary of parentage allocations between the 262-SNPs, 195-SNPs and 40-SNP panels

	262 panel				195 panel				40 panel			
	No	‡Loci comp	#Mism	LOD scores	No	Loci comp	Mism	LOD scores	No	Loci comp	Mism	LOD scores
Families not allocated	10	125 - 261	12 – 60 (Pair) 24 – 83 (Trio)	All negative	10	94 – 194	11 – 31 (Pair) 17 – 61 (Trio)	All negative	6	8 – 38	1 (Pair) 1 – 7 (Trio)	All negative

§Alloc at *	30	102 - 260	0 – 6 (Pair) 2 – 19 (Trio)	All positive	30	62 – 194	0 - 5 (Pair) 0 – 12 (Trio)	All but one family positive	19	13 – 35	0 – 1 (Pair and trio)	All positive
flAlloc at +	2	30 and 71	2 - 7 (pair) 5 and 9 (Trio)	-1.40 and -9.67	2	44 and 51	1 (Pair) 4 and 5 (Trio)	Pair positive Trio negative	14 (Pair) 5 (Trio)	20 – 34	0 – 1 (Pair) 0 – 3 (Trio)	Negative and positive

‡ Loci comp: Number of loci compared between chick and parent.
# Mism: Genotypic mismatches between chick and one parent (pair) or two parents (trio).
§ Alloc at *: Parentage allocation made at the 95% strict confidence interval.
fl Alloc at +: Parentage allocation made at the 80% relaxed confidence interval.

Table 2 shows that as the number of SNPs were reduced, the number of genotypic mismatches reduced, leading to a change in parentage allocation. The breeder's records of ten of the 43 (23.2%) families were incorrect as the suggested parents were not allocated as true parents during the 262-SNP and 195-SNP panels analyses. This also highlights the importance of this study. This number decreased to six families using the 40-SNP panel. The allocations of two families changed from the 80% relaxed confidence level during the 252-SNP panel analysis to the strict 95% level applying the 195-SNP panel. One family's allocation remained unchanged but the LOD score changed from positive (262-SNP panel) to negative (195-SNP panel).

Studies by Kaiser *et al.* (2016) (black throated blue warbler, *Setophaga caerulescens*), Liu *et al.* (2016) (rainbow trout *Oncorhynchus mykiss*) and Weinman *et al.* (2015) (superb starlings, *Lamprotornis superbus*) reported between 36 and 95 SNPs with H_O and MAF values in excess of 0.3 to confirm parentage. All three studies found that more than 90% of all relationships were correctly allocated and that more heterozygous SNPs increased the accuracy of allocations. None of these studies has taken the effect of genotyping errors on P_E into account as a high genotypic error rate will lead to more genotypic mismatches between putative parents and offspring (Heaton *et al.*, 2014). This was illustrated during the 40-SNP panel analysis, where the number of SNPs were reduced and the number of mismatches reduced due to fewer SNPs being compared. This resulted in false positive allocations of parents due to a higher LOD score.

The number of SNPs required in a panel is strongly influenced by the panel's mean expected heterozygosity and the individual SNP's H_O (Morin *et al.*, 2004; Weinman *et*

al., 2015; Kaiser *et al*., 2017) and MAF values (Strucken *et al*., 2016). Liu *et al*. (2016) concluded that a larger number of SNPs reduces the panel's sensitivity to MAF values and genotyping errors. This is stressed by Strucken *et al*. (2016) who reports that the number of markers, rather than the quality of markers dictates successful allocations, and that at least 200 SNPs should be included when one parent's genotype is known. The exclusion probability formula doesn't take MAF, H_O and genotypic error into account. Therefore, if the number of markers (and ultimately the number of mismatches) are reduced, the allocations could lead to false negative or false positive allocations. In this study the use of 40 highly informative SNPs (MAF and H_O >0.3) had less exclusion power and led to false positive allocations, compared to the 195-SNP panel with H_E and MAF values >0.1. This was confirmed in two families where only 70 (using the 262 SNP panel) and 30 (using the 195 SNP panel) SNPs could be amplified and compared. The families were allocated at the 95% and 80% confidence intervals, respectively, despite having negative LOD scores, indicating false positive allocations.

Lovebird breeders often mate close relatives to ensure that recessive colour alleles are inherited resulting in high inbreeding levels. A sufficient number of SNPs should be included in the parentage panel to accommodate this. The mean heterozygosity levels as well as the parentage allocations were similar in the 262 and 195-SNP panels. A reduced panel is more economical to genotype, and it is expected that the 195-SNP panel will be robust enough to exclude all non-parents in all domesticated lovebird families and species.

Acknowledgments:

This study was funded by the Technology Innovation Agency, SABDI grant (407/01 SABDI 16/1016). We also would like to thank all the lovebird breeders who participated in this study.

Availability of data

All SNP data is available on request from the authors. Please contact agapornis.genome.study@gmail.com to request data.

Submission numbers of parent's data: NCBI SRA accession number PRJNA355979.

Reference genome: NDXB01000000

References

Dilger W. (1960) The comparative ethology of the African parrot genus Agapornis. *International Journal of Behavioral Biology Ethology* 17, 649-685.

Eberhard J.R. (1998) Evolution of Nest-Building Behavior in Agapornis Parrots. *The Auk* 115, 455-464.

Forshaw J.M. (1989) 2nd Ed. Parrots of the World. Princeton University Press, USA.

Heaton M.P., Leymaster K. A., Kalbfleisch T. S., *et al.* and the International Sheep Genomics Consortium. (2014) SNPs for Parentage Testing and traceability in globally diverse breeds of sheep. *Plos One* 9, e94851.

Kaiser S.A., Taylor S.A., Chen N., Sillett T.S. *et al.* (2017) A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources* 17, 183-193.

Liu S., Palti Y., Gao G, *et al.* (2016) Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture* 452, 178-182.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.

McKenna A., Hanna M., Banks E. *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20, 1297-1303.

Morin P.A., Luikart G., Wayne R.K. & The SNP working group. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19, 208-216.

Robinson J.T., Thorvaldsdóttir H., Winckler W., *et al.* (2011) Integrative Genomics Viewer. *Nature Biotechnology* 29, 24-26.

Strucken E.M., Lee S.H., Lee H.K., *et al.* (2016) *Journal Animal Breeding and Genetics* 133, 13-23.

Talenti, A., Nicolazzi, E.L., Chessa, S., *et al.* (2016) *J. Dairy Sci.* 99: 3646-3653.

Van den Abeele, D. (2016) Lovebirds Compendium. About Pets Publishers, The Netherlands.

Van der Zwan H., van der Westhuizen F.H., Visser C. & van der Sluis R. (2018) Draft de novo genome sequence of Agapornis roseicollis for application in avian breeding. *Animal Biotechnology* 29, 241-246.

Van der Zwan H., Visser C., van der Sluis R. (2019) Plumage colour variations in the *Agapornis* genus: a review. *Ostrich* 90, 1-10.

Weinman L.R., Solomon J.W. & Rubenstein D.R. (2015) A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*. 15, 502 – 511.

Supporting information

Supplementary file 1: NGS sequencing, SNP discovery initial SNP panel selection, samples genotyped and construction of SNP-based parentage verification panel.

Supplementary file 2: MAF and H_O values of each SNP.

Supplementary file 3: Suggested relationships and parentage verification results.

Table 1: The Mean Expected Heterozygosity values and Exclusion probabilities of the three panels.

	SNP Panel size		
Exclusion probability	262	195	40
First parent (P_{E1})	0,999999990	0,99999980	0,9931058
Second parent (P_{E2})	0.9999999999	0.9999999999	0,99969307
Parent pair (P_{EP})	0.9999999999	0.9999999999	0,99999753
Mean H_E	0.34	0.37	0.48

Table 2: Summary of parentage allocations between the 262-SNPs, 195-SNPs and 40-SNP panels

	262 panel				195 panel				40 panel			
	No	‡Loci comp	#Mism	LOD scores	No	Loci comp	Mism	LOD scores	No	Loci comp	Mism	LOD scores
Families not allocated	10	125 - 261	12 – 60 (Pair) 24 – 83 (Trio)	All negative	10	94 – 194	11 – 31 (Pair) 17 – 61 (Trio)	All negative	6	8 – 38	1 (Pair) 1 – 7 (Trio)	All negative
§Alloc at *	30	102 - 260	0 – 6 (Pair) 2 – 19 (Trio)	All positive	30	62 – 194	0 - 5 (Pair) 0 – 12 (Trio)	All but one family positive	19	13 – 35	0 – 1 (Pair and trio)	All positive
flAlloc at +	2	30 and 71	2 - 7 (pair) 5 and 9 (Trio)	-1.40 and -9.67	2	44 and 51	1 (Pair) 4 and 5 (Trio)	Pair positive Trio negative	14 (Pair) 5 (Trio)	20 – 34	0 – 1 (Pair) 0 – 3 (Trio)	Negative and positive

‡ Loci comp: Number of loci compared between chick and parent.
# Mism: Genotypic mismatches between chick and one parent (pair) or two parents (trio).
§ Alloc at *: Parentage allocation made at the 95% strict confidence interval.
fl Alloc at +: Parentage allocation made at the 80% relaxed confidence interval.