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. 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. This was 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. 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 e*t 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 full description of the materials and methods followed are given in Supplementary File 1 and a short overview is given here. 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 1 667 639 SNPs was 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. 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. The reference genome individual's father's genotypes were used as the reference at each SNP. Due to a technical error during the analysis, only 262 of the 480 SNPs amplified for this individual. Three panels consisting of 262 SNPs (all SNPs that amplified for the father), 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. Exclusion probabilities were calculated using Cervus 3.0 applying the formulae by Jamieson & Taylor, 1997.

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 262-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.* (2017) (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. However, more research is needed in the other species before a commercial panel is compiled.

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

Data are available for reproduction of our conclusions after signing of a Material Transfer Agreement. Please contact agapornis.genome.study@gmail.com

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.

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

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.

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

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.