Supplementary File 1
NGS sequencing, SNP discovery, initial SNP panel selection

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. Sequencing of the parents yielded 44 923 Mbp of data for the father and 69 001 Mbp for the mother (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 Table S1.1.

Table S1.1: Command line arguments used during variant discovery

Action	Argument
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

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 S1.2.

Table S1.2: 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.

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. In order to identify SNPs that were found in both parents' genomes, a combined genotype file was created. A total of 1 667 629 SNPs shared between the parents were discovered. In Table S1.3 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 S1.3: 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

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 using the 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). SNPs with QD scores between 11.5 and 12.5 amounted to 4 459, from where a subset of 480 SNPs were included. The 480 SNPs with the highest QUAL scores were selected.

Lovebirds genotyped at 480 SNPs

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. 43 families from five different species and different aviaries, with full pedigree data as received from the breeders, were randomly selected to verify the exclusion power of the SNPs. These families consisted of chicks with either one parent (mother or father) or both parents. The number of samples per species, number of families for each species and the different familial relationships are presented in Table S1.4.

Table S1.4: 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.

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. 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 it was expected from the NGS sequencing and SNP selection that he should be heterozygous at each SNP.

Construction of three SNP-based parentage verification panels

All 960 lovebird samples were genotyped at the 480 selected SNPs. The father of the individual used as the reference genome 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-Weinberg 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. In Table S1.5 the mean heterozygosity values of the total population (960 birds) as well as per species for the 262-SNP 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 S1.5: 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. 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.

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 2. 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).