Comments on the manuscript (version 4-4-08) by Pedall, Gonzalez, Sauer-Gürth & Wink) "Genetic analysis of captive Lesser White-fronted Geese (*Anser erythropus*) in Germany". Reply from Wink to Ruokonen (dated 3.4.08) has been taken into consideration.

Mitochondrial DNA

In the Table below are the mtDNA haplotype frequencies from the Table 3 in Pedall *et al*. reorganized by the species and populations. Strikingly, the captive and wild populations of the lesser white-fronted goose *Anser erythropus* do not have a single mtDNA haplotype in common. In the captive population haplotypes LWFG1-4 and LWFG6-14 are found, whereas in the wild population haplotypes LWFG5 and LWFG15-21 are present. This is a very strange finding and the explanation is not obvious. Even if the sample size for the wild lesser white-fronted goose population is small, it is typical that common haplotypes are sampled with a greater probability than the rare ones, and the same fact applies to the captive populations as well. The probability that the

Table. Haplotype frequencies from Pedall *et al*. (version 4-4-08) Table 3 listed according to species and populations (as shown in Fig. 3 and in the text in Pedall *et al.*).

	greater white-fronted goose		**lesser white-fronted goose**		**bean goose**	**greylag goose**
	Russia	**Germany**	**captive**	**Russia**	**Russia**	**Germany**
GWFG	12	-	-	-	-	-
LWFG1	-	45	43	-	-	-
LWFG2	-	-	2	-	-	-
LWFG3	-	-	10	-	-	-
LWFG4	-	-	10	-	-	-
LWFG6	-	-	1	-	-	-
LWFG7	-	-	1	-	-	-
LWFG8	-	-	1	-	-	-
LWFG9	-	-	1	-	-	-
LWFG10	-	-	1	-	-	-
LWFG11	-	-	1	-	-	-
LWFG12	-	-	1	-	-	-
LWFG13	-	-	1	-	-	-
LWFG14	-	-	1	-	-	-
LWFG5	-	-	-	7	-	-
LWFG15	-	-	-	1	-	-
LWFG16	-	-	-	2	-	-
LWFG17	-	-	-	1	-	-
LWFG18	-	-	-	1	-	-
LWFG19	-	-	-	2	-	-
LWFG20	-	-	-	2	-	-
LWFG21	-	-	-	1	-	-
BG	-	-	-	-	6	-
GLG	-	-	-	-	-	5
N	12	45	74	17	6	5

captive population carries 13 lesser white-fronted goose haplotypes that are currently extinct in the wild population is extremely small, especially as it is known from previous work (Ruokonen *et al*. 2004) that there are two very common haplotypes (found in 64% of the individuals) present in the wild population, also in any wild Russian population including the one that has been sampled in Pedall *et al*.

The finding has consequences for the interpretation of the results. The purpose here would be to examine the genetic composition of the captive stocks based on the data obtained from the wild populations as a reference sample. So, now the results tell that in the captive population there are four unknown haplotypes (LWFG2-4 and LWFG6-14) and one haplotype (LWFG1) in common with the greater white-fronted goose *A. albifrons*, the latter of which could suggest that 58% of the German captive lesser white-fronted geese have a hybrid origin. Also, as seen from the Fig. 3 in Pedall *et al*., the species do not cluster into monophyletic groups and e.g. the bean goose *A. fabalis* and the greylag goose *A. anser* are more closely related to the "lineage II" than the lineage I and II are to each other suggesting that not enough resolution has been obtained with the sequences. Therefore, it is impossible to say, or even to guess, based on the tree topology, to which species some of the haplotypes belong. This is a species-level problem, as is hybridization, and not about identification of the original populations from where the captives come from, as mentioned in Wink's comments (comments from Wink to Ruokonen, dated 3.4.08).

All this suggested to us that the methodological part of the work should be reassessed. In Pedall *et al*. the PCR amplification of the cytochrome *b* gene was carried out with primers (mtA1 and mtFSH, Dietzen *et al*. 2003) originally designed for the European robin *Erithacus rubecula*, a species very distantly related to the lesser and greater white-fronted goose. It is known that mitochondrial DNA can transfer into the nucleus and integrate as part of a species nuclear genome, and this is known to have taken place in geese too (e.g. snow goose *A. caerulescens*; Quinn, 1992, 1997, *Anser* geese: Ruokonen *et al*. 2000). Especially when using PCR-primers designed for a distantly related species or so-called universal primers, the risk of PCR amplifying nuclear copies of mtDNA (numts) instead of mtDNA is pronounced. If this happens, the sequences obtained do not reflect the true phylogenetic relationships of the individuals/populations/species (e.g. Sorenson & Quinn, 1998). This is because numts reflect ancestral relationships and evolve under no functional constraints. Often, it is difficult to tell numts and mtDNA sequences apart and therefore some precautions have to be taken already in advance (see below). For example, frameshift mutations and stop codons are not necessarily present in nuclear copies simply because they have moved to the nucleus recently, and the mutation rate is 10-20 times slower in the nuclear compartment compared to mitochondria. Therefore, also the pattern typical for coding regions with substitutions taking place in 3rd codon positions can be preserved for a long period. The lack of this kind of changes in

their sequences was the argument used by Pedall *et al*. to suggest that they were of mitochondrial origin.

We designed a new pair of primers (NCL and NCH) to exactly the same nucleotide positions as in Pedall *et al*., but using the sequence of the greater white-fronted goose (complete mtDNA sequence available in the GenBank, Acc. no NC_004539, original publication: Slack *et al*. 2003, the goose collected from Scania, Sweden). We chose three wild and two captive lesser white-fronted geese, as well as three greater white-fronted geese, amplified cytochrome *b* from each individual using both primer pairs and compared the sequences obtained. All the laboratory work was carried out by a technician at the Department of Biological and Environmental Sciences, University of Helsinki.

The sequences were aligned with some sequences from Pedall *et al*. and a NJ-tree (see below) was constructed simply to illustrate the differences and similarities between the sequences. As can be seen from the tree, different sequences were obtained with different primer pairs for each individual (with the exception of A5 and M31, differing only with respect of heteroplasmic nucleotide positions), indicating that nuclear copies are present. Overall, the sequences from the same individual differed by 1 to 12 substitutions. For example, three different haplotype sequences were obtained for a wild lesser white-fronted goose A9 depending on the primers and tissues used. When compared to the haplotypes in Pedall *et al*. (haplotype names according to version distributed in 24-12-07), one of them was identical to GWFG, another identical to LWFG7 and the third differed from LWFG9 by one nucleotide substitution. Clearly then, at least two of the sequences are nuclear copies. Because of methodological reasons, one could hypothesize that the greater white-fronted goose cytochrome *b* sequence from the complete mtDNA sequence (mtDNA isolated and cloned, Slack *et al*. 2003) is of mitochondrial origin. In the same group there are also the GWFG haplotype of Pedall *et al*. found in Russian wild greater white-fronted geese, as well as a wild greater white-fronted goose from Sweden (AA9, primers NCL and NCH, DNA isolated from muscle tissue) and a wild lesser white-fronted goose from Kazahstan (A9, primers NCL and NCH, mtDNA enriched isolate). Based on the previously published mtDNA control region sequences (Ruokonen *et al*. 2000), one would expect to see very few or no differences in cytochrome *b* gene between the two species (substitution rate 5-10 times slower in cytochrome *b* compared to control region in which the differentiation between the species is approximately 1.5%; Ruokonen *et al*. 2000). The probability of amplifying the actual mtDNA should be higher with the newly designed greater white-fronted goose primers NCL and NCH compared to the robin primers mtA1 and mtFSH. However, it seems that the primers NCL and NCH do not always amplify mtDNA and this suggests that the primer location is very conserved, in fact too conserved to be used when numts are

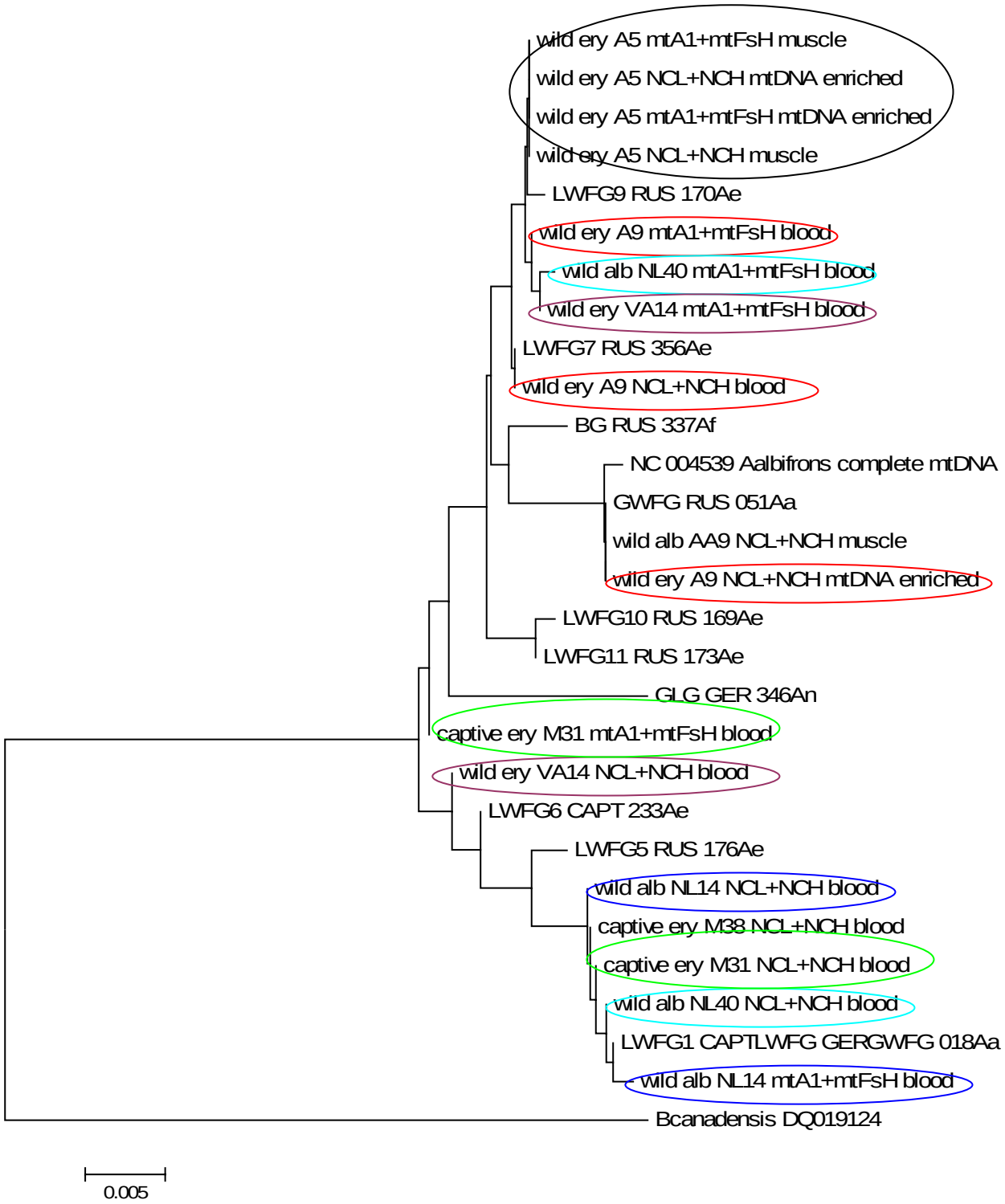

Fig. NJ-tree based on *p*-distances of the sequences from the primer comparison experiment and some of the sequences from Pedall *et al*. (sequence names in capital letters, haplotype names follow the first version of Pedall *et al*. distributed 24-12-07). Key to the sequence names: e.g. “wild ery A5 mtA1+mtFSH muscle” refers to a wild lesser white-fronted goose individual A5 PCR amplified with primers mtA1 and mtFSH with DNA isolated from a muscle tissue. Colours indicate different sequences obtained from the same individual.

a problem. In reality, several precautions should be taken to assure the mtDNA origin (e.g. DNA isolated from different tissues with differing ratios of mitochondrial and nuclear DNA, different kinds of primers, mtDNA isolation, XL PCR; see e.g. Sorenson and Quinn 1998, Ruokonen *et al*. 2000). Taking these precautions is a project which can take up to many months to accomplish.

Nuclear DNA

A list of alleles (or allele frequencies in different species/populations) shared by or private to the species/populations would have helped to evaluate the performance of Structure analysis.

Implications

The results of Pedall *et al*. do not differ from previous results in such a way that the common decision to not to use the present old captive stocks for reintroduction/population supplementation should be reconsidered.

Due to a methodological flaw, the mtDNA data in Pedall *et al*. cannot be used with any confidence to evaluate the genetic composition of the captive stocks. This analysis should be redone after the problems with nuclear copies of mtDNA are solved. Based on nuclear microsatellite data, it seems that the hybrids come from several different farms (Pedall *et al*. version 4-4-08), and this implies that there are probably also other birds affected, as the hybrids do not reproduce by themselves. This means that after removing the hybrids found, the captive stock can not be considered pure, contrary to the conclusion by Pedall *et al*.

Other comments:

Introduction: reference missing for the “old migration route” involving Germany.

References:

Dietzen C, Witt H-H, Wink M (2003) The phylogeographic differentiation of the European robin *Erithacus rubecula* on the Canary Islands revealed by mitochondrial DNA sequence data and morphometrics: evidence for a new robin taxon on Gran Canaria? Avian Sci. 3: 115–131.

Quinn TW (1992) The genetic legacy of mother goose-phylogeographical patterns of Lesser Snow Goose *Chen caerulescens* maternal lineages. Molecular Ecology 1:105-117.

Quinn TW (1997) Molecular evolution of the mitochondrial genome. Pages 3-28 in Avian molecular evolution and systematics (D. P. Mindell, Ed.). Academic Press, New York.

Ruokonen M, Kvist L, Lumme J (2000) Close relatedness between mitochondrial DNA from seven *Anser* goose species. Journal of Evolutionary Biology 13: 532–540.

Slack K, Janke A, Penny D, Arnason U (2003) Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene 302: 43-52.

Sorenson MD, Quinn TW (1998) Numts: A challenge for avian systematics and population biology. The Auk 115:214-221.

23.4.2008

Minna Ruokonen
Department of Biology
University of Oulu