UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Phylogeny and phylogeography of Atlantic Islands' *Columba* species

Ana Catarina Gonçalves Dourado

Mestrado em Biologia Humana e Ambiente

2011

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Phylogeny and phylogeography of Atlantic Islands' *Columba* species

Ana Catarina Gonçalves Dourado

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Prof^{a}. Doutora Deodália Dias
Prof. Doutor Octávio Paulo

2011

Nota prévia

A presente Tese de Mestrado está escrita na língua inglesa para ser posteriormente adaptada a uma publicação científica numa revista internacional.

Molecular Ecology foi a revista escolhida como modelo para a elaboração das referências bibliográficas por ser uma conceituada revista da área temática em que a presente Tese se insere.

Agradecimentos

Parece que é habitual agradecer às pessoas que de alguma forma contribuíram para que uma tese pudesse ser efectuada. Não fugirei à regra!

Pela proposta do tema desta tese, incentivo, orientação e apoio durante estes longos meses (mais do que inicialmente desejado!), agradeço à Professora Deodália Dias e ao Professor Octávio Paulo.

Inês Silva e Pedro Didelet Pereira: foram vocês que me deram a conhecer as rotinas laboratoriais quando ainda era voluntária (juntamente com a *Mariaaa*), e por isso vos agradeço.

Agradeço ao Dr. Paulo Oliveira por me ter fornecido algumas amostras e à Joana Gomes, que de uma forma tão prestável as entregou pessoalmente. Agradeço também ao Professor Godfrey Hewitt, com quem brevemente pude discutir os resultados desta tese.

Há pessoas a quem eu não me posso esquecer de agradecer: colegas dos grupos DDias e CoBiG^2! A todos vós, OBRIGADO! E não agradeço apenas os conselhos, a ajuda quando o computador (dizia eu!) não funcionava, mas sobretudo a AMIZADE. Tu, Joana Costa, a certa altura *já nem me devias poder ouvir*, a queixar-me das dificuldades no laboratório, da análise dos dados… e quanto à escrita da tese isso então nem se fala! Mas não eras a única, uma vez que também vocês, Ana Margarida Lopes, Ana Sofia Rodrigues, Bruno Vieira, Eduardo Marabuto, Eva Rolo, Francisco Pina-Martins, Raquel Vaz (a nossa *aquisição* da BD), Rita Oliveira Dario e Sara Silva me iam ajudando com as usuais dificuldades de uma tese. Mas também não esqueço a amizade e companhia que a Ana Vieira, Diogo Silva, Inês Modesto, Patrícia Brás, Renata Martins, Sofia Seabra e Vera Nunes me proporcionaram!

Eduardo, Joana, Bruno, Sofia, Vera e Rita: pela paciência, disponibilidade e preciosas sugestões durante a escrita da tese, vocês merecem um agradecimento especial. MUITO OBRIGADO!

Aos colegas de laboratório que me fizeram companhia ao som da Rádio Comercial, principalmente nos fins-de-semana e feriados em que pessoas *ditas normais* e *com vida social* estariam em qualquer outro lugar, obrigado. Os fins-de-semana nunca mais serão os mesmos sem uma ida à Faculdade para fazer uns quantos *PCRs* e electroforeses!

Alguns amigos fora do contexto da Faculdade foram fundamentais, visto que uma tese não se constrói apenas dentro das paredes do laboratório ou da tão *fria de Inverno e quente de Verão* sala de bolseiros! A ti, Ana Lúcia Pereira, muito obrigado por me ires *aturando*, e me obrigares a distrair quando a vontade não era essa. Também vocês, Filipa Nunes, Lese Costa, Lucas Coito, Maria Tavares e Rui Nunes foram importantes nas longas conversas ao telefone, nos divertidos jantares, nos passeios pela cidade e nos cafés no *sítio do costume*. E claro que nunca se esqueciam de fazer a tão famosa pergunta sempre que me viam: *Então, como vai essa tese?!?*

E a vocês, Pai e Mãe, nem sei bem o que dizer… Apesar de estarem no fim desta lista foram provavelmente as pessoas que mais contribuíram para que esta tese pudesse ser concluída. Pelas horas tardias para jantar, pelos fins-de-semana em que mal vos via, pelo stress frequente nos últimos meses e por me desculparem por tudo isto e muito mais, OBRIGADO!

Aqui fica publicamente e para sempre registada a minha gratidão a todos vós. Se por acaso me esqueci de alguém (e posso garantir que é bem provável que isso aconteça!) peço desculpa… não serão certamente esquecidos na Tese de Doutoramento!!!

Resumo

Os arquipélagos têm sido essenciais para a aquisição de conhecimento acerca de processos evolutivos (Bollmer *et al.* 2006; Dietzen *et al.* 2008), pois o fluxo genético é reduzido devido à existência de barreiras geográficas, a sua fauna e flora é bem conhecida devido ao reduzido tamanho das ilhas e também porque podem possuir habitats diversos. Estes factores contribuem para que os arquipélagos tenham elevados níveis de endemismo (Emerson 2002).

As diferenças entre as populações insulares e as populações continentais são geralmente maiores em espécies com baixa taxa de dispersão, em ilhas de menores dimensões e nos arquipélagos mais distantes (Frankham 1997). O reduzido tamanho das populações e o seu isolamento em algumas ilhas pode ter como resultado a existência de linhagens distintas em ilhas próximas. (Bollmer *et al.* 2006).

A filogeografia e a filogenética são duas áreas que respondem a questões complementares, pois enquanto a primeira pretende analisar os processos evolutivos que ocorrem a nível populacional, a outra pretende determinar as relações filogenéticas entre as espécies (Brito & Edwards 2009). Para responder a estas questões podem ser utilizados diversos tipos de marcadores moleculares.

O DNA mitocondrial tem sido amplamente utilizado em estudos filogenéticos devido à sua elevada taxa de evolução (Bonilla *et al.* 2010), pelo que acumula mutações mais rapidamente do que certas regiões do DNA nuclear (Avise 2009). Tem características particulares que contribuíram para que fosse um dos marcadores moleculares mais utilizados: elevada taxa de mutação, geralmente não sofre recombinação e é normalmente herdado via materna (Brito & Edwards 2009). Como os genes mitocondriais não são independentes uns dos outros, por vezes é importante comparar dados obtidos através de múltiplas genealogias (alcançadas a partir de diferentes genes), sendo o DNA nuclear fundamental (Sota & Vogler 2003).

Quando se recorre ao DNA nuclear para inferir filogenias e relações filogeográficas é necessário que sejam considerados determinados aspectos, nomeadamente o facto de isolar haplótipos nucleares poder não ser de fácil execução, assim como detectar eventos de recombinação (Avise 2009). Com o aumento do conhecimento foi possível compreender que os genes nucleares têm menos bases parcimoniosamente informativas do que o DNA

mitocondrial, tendo por isso menos poder de resolução. Para ultrapassar este problema, começou-se a recorrer à concatenação de vários fragmentos, passando também a utilizar-se intrões e não apenas zonas codificantes do genoma (Brito & Edwards 2009). Os intrões, por serem regiões não-codificantes, acumulam bases parcimoniosamente informativas mais rapidamente do que as zonas codificantes (Creer 2007).

Combinar dados moleculares com diferentes origens pode não ser fácil, pois cada fragmento possui a sua taxa de evolução, tendo um sinal filogenético diferente. Sabe-se há muito tempo que as árvores filogenéticas de cada gene podem diferir entre si (Nichols 2001). Portanto, diferentes partes do genoma podem ter histórias evolutivas distintas (Creer 2007).

Os Columbiformes são uma ordem de aves com distribuição mundial. Uma das famílias desta ordem, a Columbidae, é actualmente constituída por mais de 300 espécies (Pereira *et al.* 2007). De acordo com a *Integrated Taxonomic Information System Database*, o género *Columba* Linnaeus, 1758 contém 35 espécies.

Nos arquipélagos atlânticos (excluindo Cabo Verde), quatro espécies endémicas existem: *Columba palumbus azorica* Hartet, 1905 (Açores), *Columba trocaz* Heineken, 1829 (Madeira) (Cabral *et al.* 2005), *Columba bollii* Godman, 1872 (Canárias) e *Columba junoniae* Hartert, 1916 (Canárias) (Gonzalez *et al.* 2009).

O pombo mais abundante e disperso na Europa é o pombo torcaz, *Columba palumbus* Linnaeus, 1758 (Bruun *et al.* 1993), podendo também ser encontrado na Ásia e em África (BBC). Enquanto as populações do norte europeu são migradoras (Bea *et al.* 2003), as do sul são essencialmente sedentárias (Elias *et al.* 1998).

Das seis subespécies que estão descritas, duas existem na Europa: *C. p. azorica* e *C. p. palumbus*. A população açoriana tem sido considerada uma subespécie distinta por se encontrar geograficamente isolada (Cabral *et al.* 2005).

Nos últimos anos têm sido obtidos dados interessantes acerca destas aves, colocando-se algumas questões sobre a existência de uma completa diferenciação entre *C. p. palumbus* e *C. p. azorica* (Abrantes 2000; Grosso 2002; Duarte 2006; Silva 2007). Dados de microssatélites sugerem uma menor variabilidade genética da população açoriana, provavelmente por se encontrar isolada num sistema insular (Abrantes 2000). Resultados contraditórios foram obtidos com genes mitocondriais, que apontam no sentido de que as populações europeia e açoriana não estão diferenciadas (Grosso 2002; Duarte 2006). Um estudo recente baseado em intrões aponta para a existência de diferenciação entre as referidas populações, assim como a existência de semelhanças entre alguns exemplares da espécie *C. p.*

azorica com *C. trocaz* e *C. bollii*, enquanto outros são filogeneticamente próximos de *C. p. palumbus* (Silva 2007).

Os resultados obtidos na presente Tese de Mestrado contribuem para esclarecer as pertinentes questões previamente mencionadas.

No total foram sequenciadas 132 amostras e 2900 pb foram analisados. Encontraram-se muitas bases variáveis, seis indels e alguns indivíduos heterozigóticos em relação ao tamanho do indel.

Em relação às diversidades haplotípica e nucleotídica, *C. p. azorica* apresenta valores mais elevados do que *C. p. palumbus* e *C. trocaz* para a maioria dos intrões analisados (GHR, RP40, TROP e β-FIB), o que é muito interessante por se tratar de uma população insular, e geralmente populações insulares têm menor diversidade genética que as populações continentais que lhes deram origem (Frankham 1997). Foram detectados eventos de recombinação em quatro intrões (GHR, TROP, β-FIB e TGF-β2).

Sete SNPs interspecíficos foram descobertos, permitindo distinguir geneticamente *C. p. palumbus* de *C. trocaz*.

Quanto à análise filogenética, as filogenias obtidas diferem entre si consoante o gene que têm como base, pelo que se percebe facilmente que os filogramas e a verdadeira árvore da espécie têm diferenças, como argumentado em Nichols (2001). No entanto, existe concordância entre alguns dos fragmentos utilizados (GHR, TROP e β-FIB). Os restantes mostram padrões diferentes, evidenciando que diferentes partes do genoma podem ter histórias evolutivas distintas (Creer 2007).

Relativamente a *C. junoniae*, esta espécie chegou às Ilhas Canárias muito antes do que *C. bollii* (Gonzalez *et. al.* 2009), e isso é corroborado pelas filogenias alcançadas no presente trabalho. As espécies *C. bollii* e *C. trocaz* são filogeneticamente próximas entre si, e também com algumas amostras de *C. p. azorica*. Algumas amostras da população açoriana são filogeneticamente semelhantes a *C. p. palumbus*. Esta última subespécie não mostra diferenciação consoante o local amostrado, estando este aspecto relacionado com o seu comportamento migrador. Deste modo, é possível verificar que a subespécie açoriana não se comporta como sendo monofilética, pois surge em clades distintos nas árvores filogenéticas.

Considerando todos os resultados conhecidos acerca destas espécies, existe uma evidente discrepância entre os dados mitocondriais e nucleares. Uma possível explicação pode ser a dispersão preferencial das fêmeas (*female-biased dispersal*), que se sabe ser frequente em aves (Petit & Excoffier 2009).

Com base nos resultados apresentados, o arquipélago dos Açores deve ter sido colonizado pelo mesmo ancestral de *C. trocaz* e *C. bollii*, no mesmo evento colonizador ou pouco depois, a partir dos arquipélagos da Madeira ou Canárias. Depois, terão divergido devido a isolamento geográfico. Num passado recente, indivíduos da espécie *C. p. palumbus* terão chegado ao arquipélago açoriano, tendo ocorrido cruzamentos entre os indivíduos da espécie já estabelecida e os recém-chegados. A hibridação pode explicar as diferenças entre os marcadores genéticos utilizados, pois deve ter conduzido à introgressão do DNA mitocondrial de *C. p. palumbus* nos indivíduos que previamente habitavam as ilhas, tendo sido substituído e depois prevalecido.

Esta teoria é particularmente interessante porque a introgressão é relativamente rara em animais (Zakharov *et. al.* 2009). Portanto, *C. p. azorica* tem certamente uma história evolutiva diferente e bem mais curiosa do que se julgava, sendo presumivelmente fruto de uma dupla colonização.

Palavras-chave: género *Columba*; filogenia; intrões; introgressão; múltiplas colonizações.

Abstract

In the last decade considerable research on the genetic differentiation of genus *Columba* has been raising relevant questions, particularly about the existence of a complete differentiation beteween pigeons *C. p. palumbus* and *C. p. azorica*, since incongruent results have been achieved according to the different molecular markers used so far (Abrantes 2000; Grosso 2002; Duarte 2006; Silva 2007). While microsatellite data suggest a smaller genetic variability of the insular subspecies, *C. p. azorica*, some mitochondrial genes and a nuclear intron point out to the inexistence of a true differentiation between both subspecies. The most recent study shows that some *C. p. azorica* samples are in the same clade of *C. p. palumbus* while the others cluster together with *C. trocaz* and *C. bollii*.

In the present study, the evolutionary history of the mentioned *Columba* species was studied using nuclear DNA sequences comprising 2900 bp derived from six introns, in six nuclear genes (GHR, RP40, TROP, β-FIB, TGF-β2 and IRF2). Of the 132 samples sequenced, several have SNPs, indels and length variation polymorphism. *C. p. azorica* shows higher Hd and π values than *C. p. palumbus* and *C. trocaz* for most introns analyzed, which was not expected since, generally, insular populations have less genetic diversity than mainland populations (Frankham 1997).

As expected, *C. junoniae* comes up as an outgroup to the remaining species and very distant from the sympatric *C. bollii*. This agrees with an old colonization of the Canary Islands (Gonzales *et al.* 2009). The obtained results corroborate Silva (2007) in that the most plausible hypothesis to explain the phylogeny of woodpigeons is that Azores Islands have been colonized by *C. trocaz* and *C. bollii* ancestral and then diverged due to geographic isolation. In a recent past, *C. p. palumbus* reached those islands, and admixture occurred between both species. This might be the reason why *C. trocaz*, *C. bollii* and some *C. p. azorica* individuals are so similar, while other *C. p. azorica* and *C. p. palumbus* are in the same clade. Hybridization may have conducted to introgression of mtDNA from *C. p. palumbus* into the other individuals that inhabited Azorean Islands, and mtDNA from *C. p. palumbus* prevailed.

Instead of a mere isolated woodpigeon population, *C. p. azorica* represents a much more complex and fascinating event in pigeon evolution than it was previously believed.

Key words: genus *Columba*; phylogeny; introns; introgression; multiple colonizations.

Index

Index of Figures

Index of Tables

1 Introduction

The genus *Columba* is an interesting group of birds due to their widely studied migratory behaviors and also because these birds may be potential hosts for several pathogenic agents. Besides that, they constitute a very important resource in a few European countries as game birds.

Some studies concerning this genus focus on the research of their pathogenic agents, most of them concerning domestic pigeons, due to the effects that infections have on the performance of these birds as racing pigeons or as messengers during wars in the past (Marlier & Vindevogel, 2006). Throughout the migratory movements, birds may then carry pathogens that can be transmitted between species. This is particularly important since some of them inhabit urban areas, being eventually in contact with people. Data about avian movements may be used to improve disease surveillance schemes or to adapt preventive measures (Jourdain *et al.* 2007).

Some pigeon populations are migratory. Their migration routes may occur not only over land but also over the sea, and as a consequence they can reach nearby islands (Bankovics 2001).

1.1 Insular Systems

Oceanic islands (never connected to continents) correspond to the perfect laboratory to test for evolution in the action, a simplified model from old and complex continental systems. Islands come up with a variety of shapes and sizes, share and differ upon different biotic and abiotic environments and vary much in their distance to nearby continent(s), this having a huge effect in the animals and plants that populate them (Whittaker & Fernández-Palacios 2007).

Archipelagos have been essential for knowledge acquisition concerning evolutionary processes (Bollmer *et al.* 2006; Dietzen *et al.* 2008), such as the study of gene flow, which is reduced (and very often interrupted) by oceanic barriers. This and genetic drift make up the pillars of allopatric speciation, which is thought to be essential to the high levels of endemism in oceanic archipelagos, being microcosms for evolutionary processes (Emerson 2002).

Islands are dynamic entities, where the island area and isolation from the mainland influence the immigration and the extinction rates (Donazar *et al.* 2005; Agudo *et al.* 2011). Differences between island and mainland populations are supposed to be greater for species with lower dispersal rate, and consequently a considerable number of insular populations have lower levels of genetic diversity than the corresponding mainland populations. It is also known that insular endemic species tend to have lower genetic variation that the nonendemic ones (Frankham 1997). In several cases, despite the distance between the island and the mainland, insular and mainland populations may not be totally independent, since migration can occur, particularly for species with high dispersal capacity. In order to detect these occurrences, recent studies have been integrating not only the genetic information but also satellite tracking data (Agudo *et al.* 2011).

Small population size and isolation on island systems often results in distinct lineages on near islands. More recently, phylogenetic techniques revealed a common event – mainland source populations and colonization patterns in archipelagos (Bollmer *et al.* 2006).

Macaronesia is a biogeographical region comprising the Azores, Madeira, Canary Islands and Cape Verde archipelagos, whose high rate of endemic species makes it one biodiversity hotspot (Donazar *et al.* 2005). In fact, the number of endemic species of terrestrial organisms is estimated to be about 420 in Azores, 1419 in Madeira, 3672 in the Canary Islands (Martin *et al.* 2010) and 533 in Cape Verde Archipelago (Arechavaleta *et al*. 2005).

All these islands share a volcanic origin within the mid-atlantic. Some of them are much older than others, covering a wide range of geological ages. Their emergence above the sea level is thought to have begun about 27 million years ago (Myr) in the Selvagens Islands (Madeira), followed by Canary Islands (from 20 Myr), Madeira (15 Myr) and finally Cape Verde (10 Myr) and Azores (8 Myr) (Whittaker & Fernández-Palacios 2007). Figure 1 shows each archipelago location and each island approximated geological age (Cape Verde data not shown).

Nowadays, distances from European and African mainlands to the Macaronesian Islands have huge differences. For instance, 96 Km is the distance between Fuertventura (Canary Islands) and Stafford Point (Western Sahara), and 1370 Km separate São Miguel (Azores) from Lisbon (Europe) (Whittaker & Fernández-Palacios 2007).

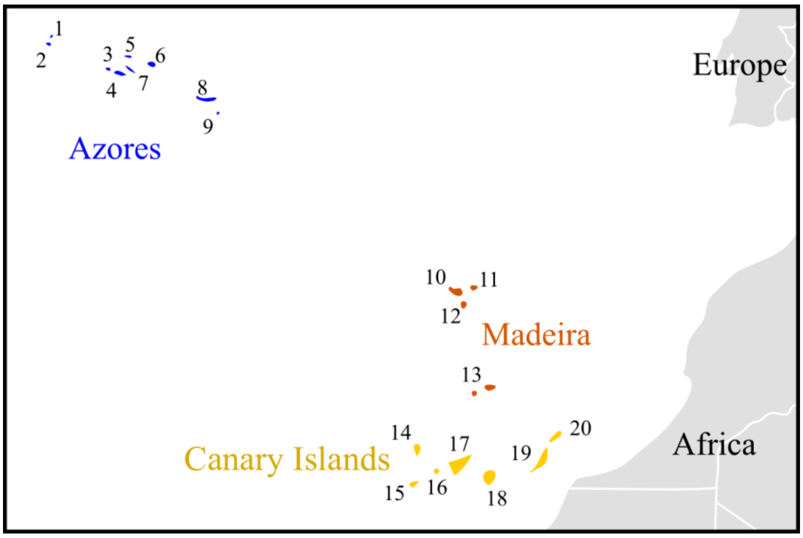

Figure 1: Part of the biogeographical area known as Macaronesia (Azores, Madeira and Canary Islands) and estimated geological age of each island (Myr): 1) Corvo – 0.71; 2) Flores – 2.16; 3) Faial – 0.73; 4) Pico – 0.25; 5) Graciosa – 2.50; 6) Terceira – 3.52; 7) S. Jorge – 0.55; 8) S. Miguel – 4.01; 9) Sta. Maria – 8.12; 10) Madeira – 4.6; 11) Porto Santo – 15; 12) Desertas Islands – 3.6; 13) Selvagens Islands – 27; 14) La Palma – 1.5; 15) El Hierro – 1.1; 16) La Gomera – 12; 17) Tenerife – 8; 18) Gran Canaria – 14; 19) Fuerteventura – 20; 20) Lanzarote – 16 (Schwarz *et al.* 2005; Whittaker & Fernández-Palacios 2007; Borges *et al.* 2009).

In spite of physically isolated, these archipelagos probably had connections in the past both between each other and with mainland areas (Fernandez-Palacios *et al.* 2011) particularly in the light of: 1) new territory has been formed as a consequence of the continuing volcanic activity; 2) the existence of consolidated wind and marine current systems in this region since 3.5 – 5 Myr, which have promoted the arrival of colonizers from the adjacent continents; 3) the lowering of sea level during the Pleistocene, conducted to the emergence of sub-marine banks, serving as stepping stones. These stepping stones may have been important for birds, which have been responsible for introducing many plant species to the islands. The banks existing to the north of Madeira Islands provided connections between Madeira and the Iberian Peninsula, for example. Dispersal opportunities between Canary Islands and Madeira were also enhanced by some of these banks and by the Selvagens Islands (Whittaker & Fernández-Palacios 2007).

Molecular phylogenetics has recently elucidated many relationships between mainland taxa and Macaronesia populations. Among plants, several plant taxa seem to have colonized Madeira from the Canary Islands, and some of them have then colonized the Azores archipelago; others are thought to have colonized firstly Madeira and then the Canary Islands (Whittaker & Fernández-Palacios 2007).

1.2 Phylogenetic and Phylogeographic Studies

Phylogeography and phylogenetics may be seen as part of a continuum that crosses the species boundary. Nevertheless, these two fields of knowledge have asked different questions, exploiting different methods: while the first one pretends to analyze the evolutionary processes that happen at the population level, phylogenetics intends to determine species relationships (Brito & Edwards 2009). It is also possible to say that phylogeography has the purpose of using phylogenetic methodologies in order to clarify historical processes in a geographic context (Nielsen & Beaumont 2009), dealing within and between closely related species (Avise 2009).

In vertebrates, mitochondrial DNA (mtDNA) has been widely used in phylogenetics due to its evolutionary rate and rapid coalescence (Bonilla *et al.* 2010). Mitochondrial DNA has several features that have contributed for being selected as a molecular marker: due to an inefficient mutation repair mechanism, mtDNA has high evolutionary rates; its genes are single copy and are not usually a target of recombination; it is generally inherited through maternal line (Brito & Edwards 2009). Because it is uniparentally inherited, molecules from different families hardly ever recombine (Avise 2009).

Animal mtDNA is very compact, and this organization must be advantageous in the cell, where the smaller mitochondria may replicate faster than the bigger ones. The rapid evolution of mtDNA in most animal species is reflected in the existence of high nucleotide sequence variation, which is a requirement for phylogeographic studies. According to mtDNA characteristics, it accumulates mutations faster than usually does nuclear DNA (nDNA) (Avise 2009).

Despite the fact that mtDNA has been routinely used to infer genealogies in phylogenetic analyses of closely related species, nDNA can be extremely important in many cases. Mitochondria represent a single linkage group, which means that mitochondrial genes are not independent from each other; so it is important to compare multiple gene genealogies with potentially different evolutionary histories (Sota & Vogler 2003).

When scientists handle nDNA, some constrains must be considered: it mutates slower than does mtDNA, isolating nuclear haplotypes may be difficult and the detection of intragenic recombination must be taken into account. But it is possible to overcome these difficulties since introns evolve faster than exons, and as a consequence they are preferred to study close relationships between individuals belonging to the same or closely related species (Avise 2009).

With recent methodological developments, another feature of nDNA was uncovered: it has been noticed that a nuclear gene usually contains a small number of informative sites, and consequently shows poor phylogenetic resolution. To overcome this, scientists increased tree resolution by concatenating several genes simultaneously. Introns are now commonly used in multilocus approaches (Brito & Edwards 2009). However, concatenation of several nuclear genes may cause some problems since the sorting of each allele in different parts of genome may conduct to incorrect species trees (Bellemain *et al.* 2008).

Combining different molecular data is not as easy as it might seem, since each gene could have their own evolution rate, containing singular phylogenetic signals. Thus, combining data is challenging. Recently, the implementation of sophisticated methodologies allows data treatment more accurately (Brito & Edwards 2009). Besides this, it is known that gene trees could differ from each other (Nichols 2001). The structure of a gene tree depends on the specific individuals that have left descendents in the next generation (in the case of mtDNA) and which of the two alleles was transmitted during reproduction (in the case of nDNA for diploid organisms) (Nielsen & Beaumont 2009).

1.3 Molecular Markers

DNA markers have been used to examine an extensive range of biological issues, including species identification, population genetics and phylogenetics (Kimball *et al.* 2009). Mitochondrial DNA has been widely used in bird molecular phylogenetics, but the same hasn’t happened with nuclear genes, rather less frequently used (Johnson & Clayton 2000).

The nuclear genome is not homogeneous, consisting of coding regions, untranslated regions, intergenic regions and introns. Molecular markers have different patterns of evolution, so they can be used at several taxonomic levels and answering different types of phylogenetic questions (Kimball *et al.* 2009).

Introns are untranslated gene regions spliced out during the formation of mature RNA molecules (Creer 2007). They are no more considered junk DNA, and some of them are known for their functional importance. Therefore, some may be evolving under the influence of natural selection (Zhu *et al.* 2009). However, since they have a small functional importance, they can be considered as neutral markers. Since introns are non-coding regions, it allows them to acquire a large number of parsimony informative sites (Creer 2007).

Introns evolve faster than exons, and they consequently have a high substitution rate. Studying this part of the genome has the advantage in the easy design of primers in the flanking exons, which are conserved regions (Bonilla *et al.* 2010).

An elementary concept in molecular phylogenetics is that a single phylogeny can be obtained from the DNA fragments under study. However, nuclear fragments often experience recombination events, and these can create mosaic genes. As a result, different parts in the DNA sequence have different evolutionary histories (Creer 2007). Recombination tends to complicate a gene tree genealogy and it should not be ignored (Avise 2009). It is possible to overcome this phenomenon in two ways: one is by studying nuclear regions with a previously known recombination rate, but these regions may not provide enough variation to answer many questions; a second approach consists in the detection of the recombination events and then the incorporation of that information into models of evolution during data analyses (Zhang & Hewitt 2003; Creer 2007; Brito & Edwards 2009). Recombination rate is different for each locus, and it is influenced by several factors, such as structural characteristics of the sequence or chromosomal location of the locus under study. It is common to find regions with higher recombination rates than mutation rates (Zhang & Hewitt 2003).

Organisms can be homozygous or heterozygous for a certain intronic sequence (since there are two copies in the diploid genome). Heterozygous bases (or ambiguities) are easily recognized in the sequence chromatogram as dual peaks of approximately the same intensity occupying the same base position (Creer 2007). Organisms that are heterozygous at a particular site have two different alleles or haplotypes. To determine the correct haplotypes (also known as to "determine the phase"), experimental and statistical approaches have been developed (Brito & Edwards 2009).

Introns frequently have insertions / deletions (indels), which often comprise a considerable part of the detected polymorphisms. In this case, patterns of indels (that appears as gaps in the aligned sequences) may contain important phylogenetic information that should not be ignored in the analyses (Zhang & Hewitt 2003). Indels can influence gene structure, pre-mRNA splicing, gene expression, gene duplication and chromosomal rearrangements (Rao *et al.* 2010). A recent study on chicken genome demonstrated that indels density is highly correlated with single nucleotide polymorphisms (SNPs) density (Brandstrom & Ellegren 2007).

Another intronic trait more complex to deal with when occurring is length variation (length variation heterozygote – LVH), and in this case, the sequence chromatogram is like an apparent corruption of the sequencing reaction by the superposition of two separate sequences occupying the same frame (Creer 2007). Even nowadays, some studies do not incorporate LVH in their analyses (Yu & Zhang 2005), but others are doing so according to different approaches (Sota & Vogler 2003; Creer *et al.* 2006; Bellemain *et al.* 2008).

SNPs can be found both in coding and non-coding regions of the genome (Primmer *et al.* 2002). A SNP can be defined as a position in the genome where more than one nucleotide can be found in a given population, thus implying different sequence alternatives (alleles). To be considered a SNP, it is necessary to have a frequency of at least 1% in a given population (Brookes 1999). These genetic markers are co-dominant (Morin *et al.* 2004) and have a Mendelian heredity (Primmer *et al.* 2002). In principle, SNPs could be bi-, tri- or tetra-alellic, but they are usually considered biallelic (Brumfield *et al.* 2003).

SNPs are believed to be the most abundant type of polymorphism in most genomes. Due to its high frequency in the genome and the existence of automatic analysis systems, their use as genetic markers has been regular in evolutionary studies (Slate *et al.* 2009). Despite being a relatively recent type of genetic marker, SNPs have been very relevant in the discovery of populations' evolutionary history (Brumfield *et al.* 2003).

In bird studies, introns of these organisms tend to be highly variable, having a high rate of SNPs (Primmer *et al.* 2002).

1.3.1 Selected Molecular Markers

Considering previous arguments in favor of introns as molecular markers in phylogenetics and phylogeography, these have been elected in this study. Other reasons must be considered behind such decision: they are abundant in the genome, their lengths are convenient for amplification and their amplification by Polymerase Chain Reaction is potentially easy (Prychitko & Moore 1997).

It is also important to use not only one but several introns, since each fragment can evolve in a differently, and consequently, the phylogenetic signal may differ between genes. Using more than one gene / intron we can bypass this problem (Brito & Edwards 2009).

1.4 Genus *Columba* - State of the Art

1.4.1 Taxonomy and Distribution of the Species in Study

Columbiformes is a very well recognized avian order worldwide. A recent study (Pereira *et al.* 2007) points to a Gondwanaland origin (as previously suggested by other studies), implying a diversification by vicariance due to continental drift. This order is subdivided in two families: Raphidae and Columbidae, being the latter represented nowadays by more than 300 living species of doves and pigeons.

According to the Integrated Taxonomic Information System Database, the genus *Columba* Linnaeus, 1758 comprises 35 species. One of the most common in Europe is the Woodpigeon, *Columba palumbus* Linnaeus, 1758 (Bruun *et al.* 1993).

In the Atlantic Archipelagos (Cape Verde not included), four endemic taxa exists: *Columba palumbus azorica* Hartert, 1905 in Azores (Portugal) *Columba trocaz* Heineken, 1829 in Madeira (Portugal) (Cabral *et al.* 2005), while both *Columba bollii* Godman, 1872 and *Columba junoniae* Hartert, 1916 are endemic to the Canary Islands (Spain) (Gonzalez *et al.* 2009).

Columba palumbus, is the most abundant and widespread of European pigeons (Bruun *et al.* 1993), having a wide distribution in Europe, Asia and Africa (BBC). In Europe, this species experienced a substantial expansion since the nineteenth century, with occasional records of nesting in Iceland (Cabral *et al.* 2005). While the northern Europe populations are described as migratory (Bea *et al.* 2003), the southern populations are mainly sedentary (Elias *et al.* 1998).

From the six described subspecies, two can be found in Europe: *Columba palumbus palumbus*, Linnaeus, 1758 (throughout Europe) and *Columba palumbus azorica* Hartert, 1905 (endemic to the Azores archipelago) (Figures 2 and 3). *Columba palumbus azorica* can be found in all Azorean islands. Azorean populations of *Columba palumbus* have been classified as distinct from the others because of their geographic isolation and slightly different morphology.

In Madeira archipelago (Portugal) existed in times the subspecies *madeirensis* (*Columba palumbus madeirensis* Tschusi, 1904), now considered extinct (Cabral *et al.* 2005).

Figure 2: *Columba palumbus palumbus* (Photo: Eduardo Marabuto).

Figure 3: *Columba palumbus azorica* (Photo: J. Palma).

Columba trocaz Heineken, 1829, commonly known as Trocaz Pigeon (Figure 4), is an endemic species from Madeira archipelago (Cabral *et al.* 2005). Despite having already existed in the island of Porto Santo, it is currently restricted to the Madeira Island (Correia-Fagundes & Romano 2011). Two factors were critical for this population decline: habitat reduction and fragmentation, as well as hunting and poisoning by humans because of the damage they often cause in agricultural fields. Due to population decline, *C. trocaz* has been considered a vulnerable species (Cabral *et al.* 2005). However, nowadays the population of this species has achieved considerable levels of recovery, which has changed the conservation status. Based on these aspects, *C. trocaz* is now officially a game bird (Paulo Oliveira, personal communication).

The Laurel Pigeon (*Columba junoniae* Hartert, 1916) is an endemic pigeon that inhabits the Canary Islands (Figure 5). Despite several constraints that occurred in the past, such as the decrease in extent of habitats, hunting and introduced predators, *C. junoniae* is not a threatened species nowadays. Evidences based on nuclear and mitochondrial genes indicate that the colonization of the archipelago by this species occurred between 5 and 20 million years ago, depending on the inference model (Gonzalez *et al.* 2009).

Figure 4: *Columba trocaz* (Photo: Carlos Cabral).

Figure 5: *Columba junoniae* (Photo: Philipp Lehmann).

As the Laurel Pigeon, the Bolles's Pigeon (*Columba bollii* Godman, 1872) is an endemic pigeon from the Canary Islands (Figure 6). Besides having been under same constrains that have affected *C. junoniae* in the last years, *C. bollii* has also recovered in terms of population size. Nowadays, both *C. junoniae* and *C. bollii* are restricted to the islands of Tenerife, El Hierro, La Gomera and La Palma, where densest laurel forest remains. It is thought that its arrival to the Archipelago occurred approximately between 1.6 and 5 million years ago, depending on the considered model. A study highlights the closer phylogenetic relationship between *C. bollii* and *C. palumbus*, suggesting a palearctic origin of *C. bollii*, later than the Laurel Pigeon (Gonzalez *et al.* 2009).

Columba livia Gmelin, 1789, the Rock Pigeon (Figure 7), has a long domestication and human associated history. Worldwide however, many birds have returned to the wild, becoming feral populations. This pigeon, considering the feral and the wild populations, has a wide distribution area (Hume 2002), being present in all continents with the exception of Antarctica. The truly wild populations occur is a more restricted area, being almost inexistent in South and North America, Asia, Australia and North of Europe. It is present in Azores and Madeira, being considered a distinct subspecies – *Columba livia atlantis* (Cabral *et al.* 2005).

Figure 6: *Columba bollii* (Photo: A. Matín).

Figure 7: *Columba livia* (Photo: J. M. Garg).

1.4.2 *C. p. azorica* – Pertinent Questions About This Subspecies

In the last decade considerable research on this bird species has been developed, raising relevant questions about the existence of a complete genetic differentiation between *C. p. palumbus* and *C. p. azorica* (Abrantes 2000; Grosso 2002; Duarte 2006; Silva 2007).

In 2000, Abrantes studied some microsatellites in these subspecies, and sampling included individuals from three European countries: France, Spain, Portugal and also the Azores archipelago. From the ten studied loci, only three of them proved to be polymorphic in the Azorean population. These data suggest a smaller genetic variability of *C. p. azorica* since they live in an insular system, with the consequent genetic isolation (Abrantes 2000).

Based on the mitochondrial gene cytochrome b, it has been suggested that the differentiation of the Azorean and some European populations of *C. palumbus* do not exist. The obtained pattern in the phylogenetic trees is consistent with a recent population expansion. However, it is stated that the sampling may have not been ideal and that the mtDNA marker may not have enough mutation rate to resolve relationships among these populations. The author recommended that the subspecies concept should not be applied to the Azorean population of *Columba palumbus* (Grosso 2002).

Another study achieved incongruent results on the matter, since the microsatellite data suggest an evident differentiation between *C. p. azorica* and *C. p. palumbus*, whereas mitochondrial (cytochrome b and cytochrome c oxidase I) and nuclear (β-fibrinogen – intron 7) markers point to the inexistence of a true differentiation. The study also included samples from Madeira Island (*C. trocaz*) and Canary Islands (*C. bollii* and *C. junoniae*), and they seem to be differentiated, suggesting an isolation by distance (Duarte 2006).

In 2007, Silva used another approach to clarify the taxonomic issue: SNPs. Results point to the differentiation between *C. p. azorica* and *C. p. palumbus*, as well as to the existence of similarities between *C. p. azorica*, *C. trocaz* and *C. bollii*. One hypothesis proposed by this author to explain these similarities is an initial colonization of Azores by ancestors of the current populations of *C. palumbus* from Europe, and then their divergence in the archipelago by geographic isolation. Then, this population has colonized Canary Islands, originating *C. bollii*, as well as the Madeira archipelago, originating *C. trocaz*. Later, a second colonization by the European Mainland *C. palumbus* may have occurred, and these new pigeons have formed hybrids with the local populations. Another hypothesis to explain the relation between *C. p. azorica* and *C. p. palumbus* and also with *C. trocaz*, is the occurrence of two colonizations from both populations. Those individuals have then originated hybrids. Silva (2007) also proposed other explanations, comprising convergent adaptation or reduction of background selection.

From the studies above mentioned, incongruent results have been achieved according to the molecular markers involved in each study - mtDNA or nDNA (microsatellites and SPNs). Thus, there is still a lack of knowledge on phylogeny and phylogeography concerning the genus *Columba*, namely the species living in the Atlantic Islands.

2. Thesis Aims

There are several goals to achieve in order to clarify the questions previously mentioned about the *Columba* species inhabiting the Azores, Madeira and the Canary Islands.

The first purpose is to increase the number of analyzed samples from previous studies, namely the following species: *C. p. azorica* (Azores), *C. trocaz* (Madeira), *C. bollii* and *C. junoniae* (Canary Islands) and also *C. livia* (from Azores, Madeira and Portugal).

In a previous work (Silva 2007), some samples were not analyzed for all genes studied, namely those from the Atlantic Islands and other European countries, such as Germany, Lithuania, Poland and Spain. Therefore, this goal is crucial in the current study as they may be interesting within the context of pigeon phylogeny.

As it was previously explained, when dealing with introns it is important to use a considerable number in order to obtain reliable results. Thus, another intention of the present study is to increase the number of introns analyzed.

It is also intended to use proper software for some assays, such as the haplotype reconstruction due to the presence of heterozygous sequences, as well as the detection of recombination.

With these approaches, this Master Thesis will contribute to clarify the evolutionary events behind the current distribution patterns and phylogeny among Atlantic *Columba* species under study, namely the colonization events and differentiation in the Azores.

3. Material and Methods

3.1 Sampling

Most of the samples used in the present work were collected between 1999 and 2000. Therefore, they were already present in the laboratory and conserved in 96% ethanol, at a temperature of -20ºC. The samples comprise tissue (muscle and liver) and feathers. New samples of *C. trocaz* (10 individuals) consisting of muscle were obtained and conserved as the previous ones. Were also obtained 54 samples (feathers) of *C. l. atlantis* from Azores, but they were in bad preservation conditions. Detailed collection data of samples can be consulted in Appendix I.

The map below (Figure 8) indicates the countries where the sampling was performed.

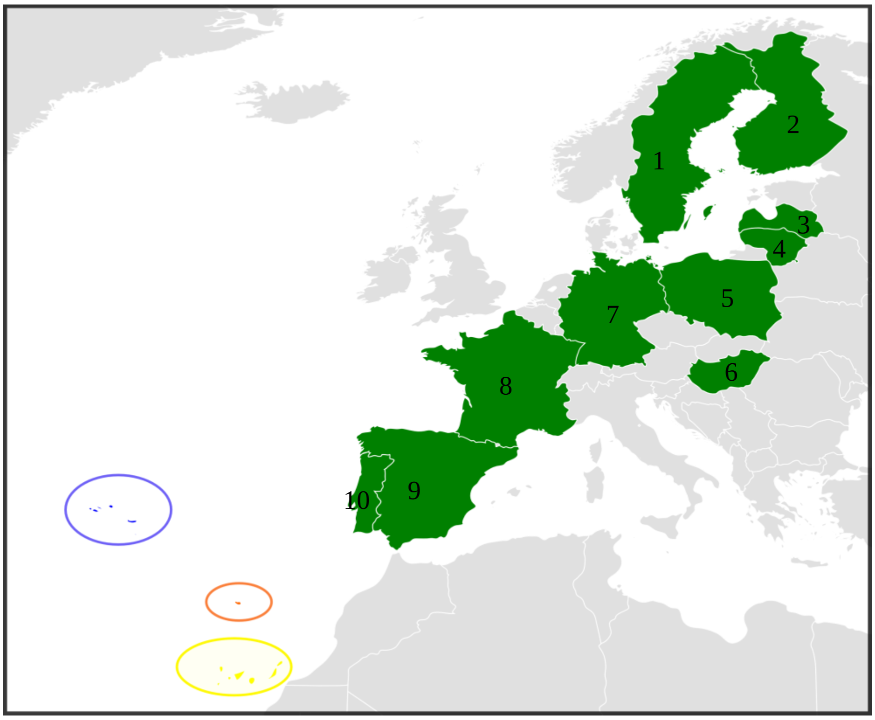

Figure 8: Sampling map of each species. Green: *C. p. palumbus* (Europe mainland – 1) Sweden, 2) Finland, 3) Latvia, 4) Lithuania, 5) Poland, 6) Hungary, 7) Germany, 8) France, 9) Spain and 10) Portugal); blue: *C. p. azorica* (Azores, Portugal); orange: *C. trocaz* (Madeira, Portugal); yellow: *C. bollii* and *C. junoniae* (Canary Islands, Spain).

3.2 Selection of Suitable Molecular Markers

Since this work is a follow up of a previous one (Silva 2007), the first goal was to improve sampling, and amplify the same introns for the new added samples, as well as for some samples that were missing in Silva (2007) work, in order to increase the work robustness (observe Appendix II). Thus, in this work, the same four introns of Silva (2007) were used, from four different genes: growth hormone receptor (GHR – intron 9); ribosomal protein 40 (RP40 – intron 5); tropomyosin (TROP – intron 5) and fibrinogen β-chain (β-FIB – intron 7).

Nevertheless, and in the same purpose of increase work robustness, three new introns were chosen for this work based on some principles: they had to be present in different chromosomes, primers should have already been designed and the amplifications should have been previously tested in several species, in order to reduce the optimization procedures. So, fragments chosen were: ATP-citrate lyase (ACLY – intron 16) (Fonseca 2006), transforming growth factor - β2 (TGF-β2 – intron 5) and interferon regulatory factor 2 (IRF2 – intron 2) (Kimball *et al.* 2009). These introns have been tested in other works, such as Prychitko *et al.*(1997), Johnson *et al.* (2001), Russello & Amato (2004), Fonseca (2006), Brumfield *et al.* (2008) and Oliveros & Moyle (2010).

GHR gene encodes a protein that is a transmembrane receptor for growth hormone. Binding of growth hormone to the receptor leads to the activation of an intra- and intercellular signal transduction pathway leading to growth (GeneCards Online Database[1]).

RP40 gene acts in ribosome formation and in the regulation of ribosome activity. It is also a precursor for a membrane associated laminin receptor. It is known that chicken has a single gene, while mammals have several copies, being some of them pseudogenes (Friesen *et al.* 1999).

TROP is a protein involved in the control of muscle fibers' contraction and relaxation (Friesen *et al.* 1999).

β-FIB is a protein with two major functions: yielding monomers that polymerize into fibrin and acting as a cofactor in platelet aggregation (GeneCards Online Database[2]).

ACLY is the first enzyme responsible for the synthesis of cytosolic acetyl-CoA in many tissues, with a simultaneous hydrolysis of ATP (GeneCards Online Database[3]).

TGF-β2 plays important functions in growth, development, inflammation, repair and host immunity (Clark & Coker 1998).

IRF2 protein has several functions, namely acting in interferon regulation (UniProt Online Database).

3.3 DNA Extraction, Amplification and Sequencing

As some samples had already been sequenced for some introns in the previous mentioned work of Silva (2007), it was not necessary to carry out the DNA extraction on these because the DNA was available and adequately preserved in the laboratory.

For the 64 sampled individuals, DNA extraction was performed following different techniques. While tissue samples were extracted with the commercial kit E.Z.N.A.® Tissue DNA Kit (Omega) following the manufacturer's instructions (with few adjustments achieve better results), DNA from feathers was extracted through standard phenol/chloroform procedures (Appendix III), since it was the most effective method available in the laboratory for this kind of samples. DNA integrity was confirmed on 1% agarose gel electrophoresis. In order to avoid DNA degradation, extractions were then frozen at -20ºC.

After DNA extraction, Polymerase Chain Reactions (PCRs) were performed in GeneAmp PCR System 2700 Thermocyclers (Applied Biosystems, USA). Table 1 summarizes primers used and information about the chromosomes where the introns are located.

Even after a great effort to amplify the ACLY intron, according to different optimizations on the PCRs reactions, it was never possible to obtain any amplification product. So, this intron was excluded.

Table 1: Genes, chromosomes, introns, primers sequence and their references.

Genes	Chromosomes	Introns	Direction	Primers sequences (5'→ 3')	References
GHR	Z	9	F	GATTAAAGGGATTGACCCAGATCTT	Silva 2007
			R	AAGTCGTCATTGTACAGCTGTG	
RP40	15	5	F	GGGCCTGATGTGGTGGATGCTGGC	Friesen *et al.*, 1999
			R	GCTTTCTCAGCAGCAGCCTGCTC	
TROP	10	5	F	GAGTTGGATCGCGCTCAGGAGCG	Friesen *et al.*, 1999
			R	CGGTCAGCCTCCTCCGCAATGTGCTT	
FIB	4	7	F	GGAGAAAACAGGACAATGACAATTCAC	Pryckitko *et al.*, 1997
			R	TCCCCCAGTAGTATCTGCCATTAGGGTT	
ACLY	27	16	F	ATCTCTACGCTCTGCTTATG	Fonseca 2006
			R	GAATGCCCTCAGCAATAATG	
TGF-β2	3	5	F	GAAGCGTGCTCTAGATGCTG	Kimball *et al.*, 2009
			R	AGGCAGCAATTATCCTGCAC	
IRF2	4	2	F	ATGTCTTTGGGTCGGGTTTA	Kimball *et al.*, 2009
			R	GAAACTGGGCAATTCACACA	

Table 2 represents PCR reagents and their concentration used in the amplification of each fragment.

Table 2: Final concentrations of reagents used in the amplification of each fragment, as well as the reactions' final volume.

	Genes					
Reagents	**GHR**	**RP40**	**TROP**	**β-FIB**	**TGF-β2**	**IRF2**
Buffer	1 x	1 x	1 x	1 x	1 x	1 x
MgCl2 (mM)	1	1	1	1.2	2	1.4
dNTPs (mM)	0.5	0.5	0.4	0.6	0.4	0.6
Primer F (pmol/µl)	0.4	0.7	0.5	0.6	0.8	1
Primer R (pmol/µl)	0.4	0.7	0.5	0.6	0.8	1
BSA (µg/µl)	0.6	0.6	0.6	0.1	0.25	0.1
Taq polymerase (U)	0.25	0.25	0.25	0.25	0.25	0.25
DNA (µl)	2.5	2.5	2	3	3	1.5
Final Volume (µl)	25	25	12.5	25	25	12.5

Each intron required a slightly different PCR cycle:

- GHR: there was an initial denaturation at 94ºC for 5 min followed by 40 cycles of 94ºC for 30 s, annealing temperature of 51ºC for 30 s, 72ºC for 30 s, and a final extension period at 72ºC for 10 min.

- RP40: amplifications were carried out by an initial denaturation at 94ºC for 5 min followed by 40 cycles of 94ºC for 30 s, annealing temperature of 58.5ºC for 30 s, 72ºC for 30s, and a final extension step at 72ºC for 10 min.
- TROP: PCR protocol consisted of an initial denaturation at 94ºC for 5 min followed by 45 cycles of 94ºC for 30 s, annealing temperature of 66ºC for 30 s, 72ºC for 30 s, and a final extension step at 72ºC for 10 min.
- β-FIB: PCRs were carried out by an initial denaturation at 94ºC for 5 min followed by 45 cycles of 94ºC for 45 s, annealing temperature of 55ºC for 1 min, 72ºC for 1 min 30 s, and a final extension step at 72ºC for 10 min.
- TGF-β2: there was an initial denaturation at 94ºC for 5 min followed by 40 cycles of 94ºC for 45 s, annealing temperature of 58ºC for 30 s, 72ºC for 45 s, and an ending extension period at 72ºC for 10 min.
- IRF2: initial denaturation at 94ºC for 5 min followed by 5 cycles of 94ºC for 30 s, annealing temperature of 56ºC for 30 s, 72ºC for 45 s, then 30 cycles of 94ºC for 30 s, annealing temperature of 57ºC for 30 s, 72ºC for 45 s, and a final extension step at 72ºC for 10 min.

Negative controls were included in all PCRs in order to confirm the absence of contaminants. Positive controls (*Gallus gallus*) were included to verify the reactions achievement, particularly during the optimization procedures.

Until further processing, PCR products were stored at 4ºC.

The amplified fragments were then ran in 1% agarose gels stained with RedSafe (iNtRON Biotechnology, Korea) at 90V for about 25 minutes, along with the molecular weight marker HyperLadder™ IV (Bioline), to confirm amplification and the expected fragments sizes. Gels were photographed under UV light in a trans-illuminator. Images were taken and treated with a Kodak EDAS 290 Electrophoresis Documentation and Analysis System (Kodak).

PCR products were then purified using the commercial kit SureClean (Bioline, UK), according to the manufacturer instructions.

Sequencing reactions were carried out in both directions with the BigDye® Terminator v3.1 Sequencing Kit (Applied Biosystems, USA), according to the manufacturer protocol, with the primers previously used for the amplification of each fragment. Reactions were performed in GeneAmp PCR System 2700 Thermocyclers (Applied Biosystems, USA). The amplification conditions consisted on an initial denaturation at 96ºC for 1 min, followed by 25 cycles of 96ºC for 10 s, annealing temperature of 50ºC for 5 s, and a final extension step at 60ºC for 4 min.

The amplified products were purified through a 70% ethanol precipitation, described next. The total reaction volume was transferred to a 1.5 ml tube with 1 µl of 3 M of sodium acetate and 25 µl of absolute ethanol. This mixture was subsequently incubated on ice for 30 min, and then centrifuged at 10.000 g for 25 min. The supernatant was discarded and 300 µl of 70% ethanol were added to each tube, which were centrifuged for 15 min at the same speed; this last step was performed a second time. Finally, the supernatant was completely discarded and the samples were air-dried in the dark, until further processing.

After a preparation of the purified product with 20 µl of formamide, followed by a denaturation process at 96ºC for 3 min, the products were sequenced in an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems, USA) available in the laboratory.

Due to their considerable length, some introns were sequenced in a specialist company (Macrogen Inc., Korea).

3.4 Molecular Analysis

Firstly, obtained sequences were edited in Sequencher v 4.0.5 (Gene Codes Corporation). Sequences were then blasted in NCBI online database (www.ncbi.nlm.nih.gov/) to guarantee the amplification of the correct fragment. Afterwards, they were edited with BioEdit v 7.0.9.0 (Hall 1999), and subsequently aligned with ClustalX v 2.0.12 (Thompson *et al.* 1997).

Haplotype phasing from heterozygous individuals was carried out with PHASE v 2.1.1 (Stephens & Donnelly 2003) for each species and intron separately according to the default parameters. Indels were not included in this determination.

After haplotype determination, two datasets are available for further analyses for each intron: while one contains the individuals with unresolved ambiguities (heterozygous sequences), the other matrix consists of phased sequences, where some individuals are represented by more than one input sequence in that dataset (*a* and *b* sequences, for each haplotype).

In the case of samples with ambiguities and length variation in the sequence (only β-FIB intron 7), a combination of possible options was determined manually (i.e., two length variations and two haplotypes are possible to combine in four different ways). Thus, those individuals are represented by four input sequences in the final data matrix (lvh1a, lvh1b, lvh2a and lvh2b). In the case of samples without ambiguous bases, only two haplotypes exist, one with the insertion and the other one without it (lvh1 and lvh2).

The number of variable sites, number of haplotypes, haplotypic diversity (Hd), nucleotide diversity (π) and the minimum number of recombination events (Rm) were estimated with DnaSP v 5.10.01 (Librado & Rozas 2009). Number of indels was determined observing and comparing sequences.

To test departure from the neutral model, neutrality tests statistics (Tajima's D and Fu's Fs tests) were determined in Arlequim v 3.5.1.2 (Excoffier & Lischer 2010), using phased data in these analyses.

3.5 Phylogenetic and Phylogeographic Analysis

The conversion of the obtained matrices for the phylogenetic and phylogeographic analyses was undertaken with the use of the software Concatenator v 1.1.0 (Pina-Martins & Paulo 2008).

Phylogenetic analyses using Maximum Parsimony for each intron (heterozygous and phased matrices independently) were performed employing PAUP v 4.0.d99 (Swofford 2002). Heuristic searches were performed with 1000 replicates and gaps were treated as a fifth base. Bootstrap support values (1000 replicates) and consensus trees were also obtained with this software.

After the phasing process, phylogenetic analyses were performed again with the new matrices, according to the mentioned parameters.

With Concatenator v 1.1.0 (Pina-Martins & Paulo 2008) it was obtained a concatenated matrix of 5 out of 6 fragments (IRF2 was not included). Whenever a sample of a given intron was not available, "N" ambiguities were added to the matrix. The analysis only included samples which had 4 or 5 introns sequenced; the others were excluded (see Appendix IV). Phylogenetic analysis using Maximum Parsimony was then performed according to the preceding parameters.

When sequences were available, *Columba livia* was the chosen outgroup; in the other cases it was used *Columba junoniae*.

Phylogeographic analysis, performed in Network v 4.6.0.0 (Bandelt *et al.* 1999), resulted in Median-Joining networks for both heterozygous and phased matrices.

The obtained sequences will be submitted to GenBank (NCBI online database).

4. Results and Discussion

4.1 Extractions and Amplifications - Laboratorial Constraints

Samples of *Columba livia atlantis* were not in good storage conditions when arrived to the laboratory, and most of them consisted of only one or two small feathers, with none or very low DNA quantity. Not even the phenol / chloroform method (Appendix III) was effective to extract the DNA. Due to the mentioned constraints, the laboratorial work with these 54 samples ended prematurely, and as a consequence they do not appear in the analyses.

Intron ACLY was not possible to amplify properly even after innumerous optimizations trials. One reason for this may be the fact that the primers used to amplify this fragment were designed based on a relatively distant species, and it may have compromised the primers connection to the target sequence in *Columba* species.

Generally, the optimal reaction conditions were not easy to achieve. It may be due to the homology between the genomes from which the primers were designed and the *Columba* genome.

After overcoming these restrictions, samples could then be amplified and sequenced for each fragment. Table 3 describes briefly the samples included further in the analyses for each intron. Complete information can be found in Appendix II.

Table 3: Number of amplified samples of each species for each fragment. N corresponds to the number of samples.

				Genes					
Species	Origin	Sample code	N	GHR	RP40	TROP	β-FIB	TGF-β2	IRF2
C. trocaz	Madeira	CtM1 - CtM15	15	15	15	13	13	15	9
C. p. azorica	Azores	CpaA20, CpaA23-CpaA27, CpaA29-CpaA37, CpaA39-CpaA42	19	19	19	9	11	17	12
C. bollii	Canary Islands	CbC1-CbC2	2	2	1	2	1	-	-
C. junoniae	Canary Islands	CjC	1	1	1	1	1	-	-
C. p. palumbus	Finland	CppFin1-CppFin3, CppFin5-CppFin8, CppFin10, CppFin13, CppFin15, CppFin16, CppFin20, CppFin21, CppFin24	14	13	7	4	1	1	1
	France	CppFr88, CppFr89, CppFr96-CppFr100, CppFr102, CppFr107, CppFr109, CppFr110, CppFr113, CppFr114	13	13	6	-	1	1	-
	Germany	CppG6, CppG7, CppG9	3	2	3	1	1	1	-
	Hungary	CppH1-CppH7, CppH10, CppH11, CppH13, CppH15, CppH23, CppH30	13	13	6	3	4	2	-
	Latvia	CppLa1-CppLa13	13	12	3	-	-	1	-
	Lithuania	CppLit1-CppLit3	3	3	2	-	-	1	-
	Poland	CppPol1, CppPol2, CppPol5	3	3	2	1	1	1	-
	Portugal	CppPt1-CppPt17	17	17	14	4	5	3	1
	Spain	CppSp58, CppSp80, CppSp85	3	3	3	1	1	1	1
	Sweden	CppSw1-CppSw8, CppSw10, CppSw14, CppSw16-CppSw18	13	13	4	1	1	1	-
		Total	132	129	86	40	41	45	24

4.2 Sequence Analysis

Considering all introns, a maximum of 2900 bp could be analyzed for a single specimen. The full length of each intron can be seen in the table below (Table 4). In the case of β-FIB intron 7, the value corresponds to the biggest sequence, since this is variable according to the presence or absence of indels. To achieve such lengths, the obtained fragments were blasted with homologous sequences on the NCBI online database, and then cut to exclude the adjacent portions of exons.

Table 4: Length of each intron and the total number of base pairs analyzed.

	Genes						
	GHR	**RP40**	**TROP**	**β-FIB**	**TGF-β2**	**IRF2**	**Total**
Length (bp)	259	237	356	855	585	608	2900

Heterozygous base positions (or ambiguities) were found in all introns studied and could be easily identified due to an overlapping curve in the chromatogram after sequencing. Double peaks have roughly half the ordinary size.

Introns β-FIB and TGF-β2 presented indels in several samples. When those indels were present in homozygosity (meaning that both copies of the intron have the same length), they could only be detected whenever sequences from different individuals were aligned.

However, some individuals have a length variation in some introns (LVH – length variation heterozygotes). This means that one copy of the intron, in one of the chromosomes, is bigger than the other. Sequences were identifiable because chromatograms developed overlapping peaks downstream from a particular base position when sequenced using each primer.

In some samples it was possible to confirm the existence of multiple indels in heterozygozity, but those chromatograms were inconclusive to infer correct sequences. The optimal procedure to overcome this situation would have been cloning those samples for that intron and later infer each haplotype, but it was not possible to accomplish. Thus, they were not included in the analyses.

In β-FIB intron 7 several indels were found, ranging from 1-6 bp: three indels were single base, one indel has 2 bases, one has 5 bases and the other has 6 bases. Some indels were found in homozygosity. If on one hand indels and LVHs were found in all species, the first indel detected in the sequence (1 bp) was only found in *C. trocaz* (all samples have the insertion) and in six *C. p. azorica* samples. The second (5 bp) was also present in all *C. trocaz* samples and in three *C. p. azorica* individuals. These indels were not found in any *C. p. palumbus* sample.

Only one 1 bp indel was found in 9 samples (2 in *C. p. palumbus* and 7 in *C. p. azorica*) in TGF-β2 intron 5. The indel was found in heterozygosity, i.e., there is length variation (LVH). To simplify the analyses, and since it was only 1 bp pair indel, it will be no further considered.

Measures of genetic diversity of each gene were determined, as well as neutrality tests and detection of indels. Full data can be observed on Table 5.

In the case of GHR, the number of obtained haplotypes is not high and is even smaller for *C. trocaz* (only two haplotypes were found). The number of variable sites range from one (*C. trocaz*) to five (*C. p. azorica*). Both Hd and π are higher in the azorean population. Two minimum recombination events were detected in this intron. Neutrality tests are not statistically significant.

Fragment RP40 shows a similar situation, where there are fewer haplotypes in *C. trocaz* and nucleotide diversity is also higher in *C. p. azorica*. Hd has similar values for *C. trocaz* and *C. p. azorica*, but *C. p. palumbus* haplotypic diversity is much smaller. No recombination events were detected in this intron. In the case of *C. trocaz*, neutrality tests show no significance.

C. p. palumbus has statistical significance values for Tajima's D test (-1.455; P = 0.031) and Fu's Fs test (-2.873; P = 0.039) tests. Under standard neutral model, Tajima's D value is expected to be near zero (Quintela *et al.* 2010). The obtained negative Tajima's D value indicates a departure from the neutral model, which might be explained with linkage to a locus under directional selection or changes in population size. Fu's Fs test is sensitive to population demographic expansion, which generally leads to negative values (Quintela *et al.* 2010). This may be the case in this species (-2.873; P = 0.039).

C. p. azorica also exhibits negative Tajima's D and Fu's Fs values (-1.566; P = 0.016 and -1.436; P = 0.037 respectively). The reasons mentioned before may explain these values.

In the case of TROP, results are basically similar regarding the number of haplotypes and the number of variable sites. Here, Hd assumes almost equal values in both *C. palumbus* subspecies, but is lower in *C. trocaz*. Once again, *C. p. azorica* shows the highest nucleotide diversity. A single Rm was detected.

A significant Fu's Fs value in *C. p. azorica* (-3.640; P = 0.002) is an indicative of demographic expansion or an excess of recent mutations (Ramos-Onsins & Rozaz 2002; Quintela *et al.* 2010).

C. trocaz has a significant Tajima's D value (-1.513; P = 0.039), which may be the result of changes in population size or linkage to a locus under directional selection (Quintela *et al.* 2010).

In the case of β-FIB intron 7, the number of haplotypes is higher in *C. p. palumbus*. The number of variable sites is almost the same for both *C. palumbus* subspecies, but is much lower in *C. trocaz*. As in GHR, RP40 and TROP fragments, nucleotide diversity is higher in the azorean pigeons. Four recombination events were detected in this intron.

Regarding the neutrality tests, both *C. p. palumbus* and *C. trocaz* have significant negative Fu's Fs values (-14.274; P = 0.000 and -5.142; P = 0.003 respectively).

For the TGF-β2 fragment, the number of haplotypes and the number of variable sites is equal for *C. p. azorica* and *C. trocaz*, but slightly higher in *C. p. palumbus*. In contrast to what happened for the 4 previous fragments, both Hd and π show lower values in *C. p. azorica*. None of the neutrality tests has statistical significance.

Intron IRF2 indicates the inexistence of variability among *C. p. palumbus* and *C. p. azorica*, while *C. trocaz* still shows a considerable haplotypic and nucleotide diversity. From the available data, this intron does not show enough resolution to infer a phylogeny since it is not variable enough in all species.

It is important to notice that for this fragment some sequences did not attain the desirable length; due to software constrains, only one haplotype was found in *C. p. palumbus* (however, the phylogram – Figure 25 – shows three different haplotypes for this subpecies).

Due to the existence of only one haplotype for *C. p. palumbus* and *C. p. azorica*, Tajima's D test result is zero – this intron is probably under a neutral model or evolving slowly (Quintela *et al.* 2010). Fu's Fs cannot be determined. In the case of *C. trocaz*, the values are not significant.

Table 5: Genetic variability at each of the six nuclear intronic regions, for each species. Number of haplotypes, number of variable sites, haplotypic diversity (Hd), nucleotide diversity (π) and minimum number of recombination events (Rm) were calculated employing DnaSP v 5.10.01. Values of standard deviations are shown for Hd and π. Neutrality tests (Tajima's D and Fu's Fs tests) were determined in the software Arlequim v 3.5.1.2. P values for Tajima's D and Fu' Fs tests are given in brackets, and * indicates significance ($P < 0.05$). Number of individuals (N) and number of indels are also shown.

Genes	**Species**	**N**	**Number of haplotypes**	**Number of variable sites**	**Hd**	***π***	**Number of indels**	**Rm**	**Tajima's D (P)**	**Fu's Fs (P)**
GHR	*C. p. palumbus*	92	5	4	0.365 ± 0.042	0.00196 ± 0.00025	0	2	- 0.427 (0.388)	- 1.906 (0.225)
	C. p. azorica	19	5	5	0.653 ± 0.054	0.00662 ± 0.00088	0		1.134 (0.889)	1.165 (0.776)
	C. trocaz	15	2	1	0.434 ± 0.070	0.00207 ± 0.00033	0		1.124 (0.886)	1.369 (0.680)
RP40	*C. p. palumbus*	50	5	5	0.264 ± 0.060	0.00153 ± 0.00042	0	0	- 1.455 * (0.031)	- 2.873 * (0.039)
	C. p. azorica	19	4	3	0.560 ± 0.059	0.00304 ± 0.00046	0		- 1.566 * (0.016)	- 1.436 * (0.037)
	C. trocaz	15	3	2	0.591 ± 0.054	0.00284 ± 0.00039	0		0.657 (0.764)	0.656 (0.615)
TROP	*C. p. palumbus*	15	7	6	0.677 ± 0.082	0.00239 ± 0.00042	0	1	- 1.368 (0.067)	- 3.640 * (0.002)
	C. p. azorica	9	4	4	0.608 ± 0.086	0.00481 ± 0.00057	0		1.147 (0.847)	1.018 (0.704)
	C. trocaz	13	3	2	0.177 ± 0.106	0.00081 ± 0.00051	0		- 1.513 * (0.039)	- 0.265 (0.175)
β-FIB	*C. p. palumbus*	15	18	18	0.934 ± 0.014	0.00470 ± 0.00039	2	4	- 0.226 (0.467)	- 14.274 * (0.000)
	C. p. azorica	11	9	20	0.829 ± 0.035	0.00747 ± 0.00036	3		- 0.917 (0.847)	0.336 (0.579)
	C. trocaz	13	7	10	0.510 ± 0.081	0.00157 ± 0.00039	5		- 1.488 (0.074)	- 5.142 * (0.003)
TGF-β2	*C. p. palumbus*	14	7	6	0.826 ± 0.00177	0.00262 ± 0.00030	1	1	- 0.301 (0.429)	- 1.799 (0.120)
	C. p. azorica	17	5	4	0.683 ± 0.049	0.00203 ± 0.00015	1		0.340 (0.726)	- 0.212 (0.456)
	C. trocaz	15	5	4	0.786 ± 0.035	0.00265 ± 0.00021	0		1.115 (0.857)	0.364 (0.619)
IRF2	*C. p. palumbus*	4	1	0	0	0	0	0	0 (1.000)	—
	C. p. azorica	12	1	0	0	0	0		0 (1.000)	—
	C. trocaz	8	3	2	0.575 ± 0.112	0.00159 ± 0.00034	0		0.944 (0.848)	0.586 (0.617)

On overall it is easy to perceive that *C. p. azorica* is a very diverse subspecies which has in general higher Hd and π values (notice the values for fragments GHR, RP40, TROP and β-FIB), with remarkably high values in comparison with the other considered taxa. This is against a generalized idea that postulates that genetic diversity in insular species is lower than in mainland populations (Frankham 1997). Neutrality tests in *C. p. azorica* only show statistical significance for RP40, pointing not only to a demographic expansion but also linkage between this locus and other under directional selection.

Tajima's D test in *C. p. palumbus*, is again only statistically significant for RP40, while regarding Fu's Fs test, three out of six fragments (RP40, TROP and β-FIB) point to the existence of demographic expansion. This could be related with the dispersal capacity shown by these migratory birds.

In the case of *C. trocaz,* the results indicate not only a slight demographic expansion but also linkage to a locus under directional selection. With the exception of IRF2, *C. trocaz* has always a smaller number of haplotypes and variable sites than *C. p. palumbus* and *C. p. azorica*, this time in agreement with Frankham (1997) for insular populations although we don't have a mainland conspecific taxon to compare with.

Several SNPs were found in all introns, all of them being bi-allelic. Some of them are interspecific, allowing the differentiation between *C. p. palumbus* and *C. trocaz* (Table 6).

Table 6: Interspecific SNPs found in the fragments GHR, TROP, β-FIB and TGF-β2.

		Species	
Genes	**SNP position**	***C. p. palumbus***	***C. trocaz***
GHR	213	C	T
	233	A	T
TROP	179	G	A
	182	A	G
β-FIB	291	G	A
TGF-β2	79	G	A
	520	A	G

4.3 Phylogenetic and Phylogeographic Analysis

Due to the short length of most fragments, and therefore a small number of parsimonious informative sites, bootstrap values are relatively low. In order to minor this phenomenon, it was opted to clarify how many trees were really obtained in each analysis (since only one is shown) and also to include a consensus tree, to verify the existence of discrepancies.

Maximum parsimony phylograms for both heterozygous and phased sequences of each intron can be consulted in the next pages. Consensus cladograms for each analysis are shown in Appendix V.

4.3.1 GHR Intron 9

According to the GHR intron 9 parsimony tree obtained before the phase determination (Figure 9), there are three groups: one is only comprised by *C. junoniae*, the earliest species inhabiting the Canary Islands; the second is mainly constituted by *C. trocaz* and *C. bollii*, and finally the third clade comprises all samples of *C. p. palumbus*. As expected from the previous work (Silva 2007), *C. p. azorica* samples are present in the two latter clades – sixteen being phylogenetically closer to *C. p. palumbus* and three of them align with *C. trocaz* and *C. bollii*. This contradicts Grosso (2002) and Duarte (2006), which found no differentiation among the subspecies *C. p. palumbus* and *C. p. azorica*, according to mtDNA data.

As expected, *C. junoniae* comes up as an outgroup to the remaining species and very distant from the sympatric *C. bollii*. This agrees with an old colonization of the Canary Islands (Gonzalez *et al.* 2009).

C. trocaz and *C. bollii* are not so distant from each other, since they appear in the same clade, sharing one haplotype. This suggests that they may have arisen from the same macaronesian colonization event from a common ancestor, probably *C. palumbus*.

In spite of being a widespread paleartic taxon, with samples from a wide range of European countries, *C. p. palumbus* shows slight differentiation. This might corroborate the idea of a widespread and migratorial population (Bea *et al.* 2003), whose vagility is preventing the species isolation and speciation.

Some minor differences are found between Figure 9 and the correspondent consensus tree (Appendix V, Figure 28), although these can be considered irrelevant for the present study. For instance, the samples CppFin10 and CppPol1 are polytomic in the parsimony tree while in the consensus they are sister groups; in the other clade, some *C. trocaz* samples have a different position.

The phased matrix (with haplotypic data) was submitted to the same maximum parsimony analysis. Figure 10 shows one out of fifty-four phylograms, which shows the same general topology as Figure 9. However, in the case of samples with two different haplotypes (represented by *a* and *b* after the sample name), they come-up separated in different positions within the tree, although still in the same clade. In the case of *C. p. azorica*, only sample CpaA20 exhibits ambiguous bases, and both phased sequences appear near *C. p. palumbus*.

There are few differences between the shown phylogram (Figure 10) and the consensus tree (Appendix V, Figure 29). In the later, some *C. p. palumbus* samples arise as sister groups instead of being part of the polytomy, and the same happens with some samples of *C. trocaz*. Nevertheless, none of these small changes modify the phylogenetic relationships.

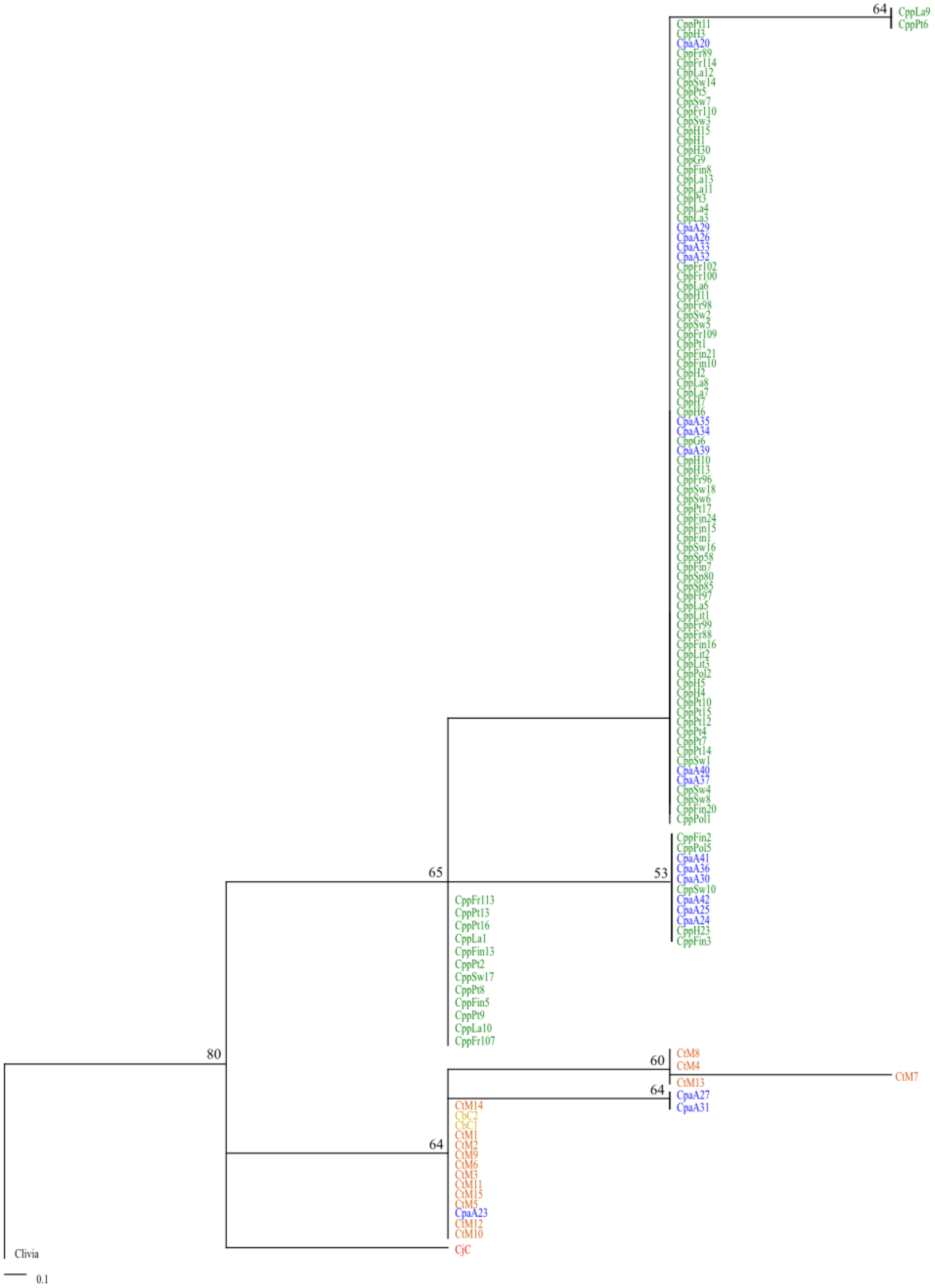

Figure 9: Maximum parsimony tree based on GHR intron 9 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

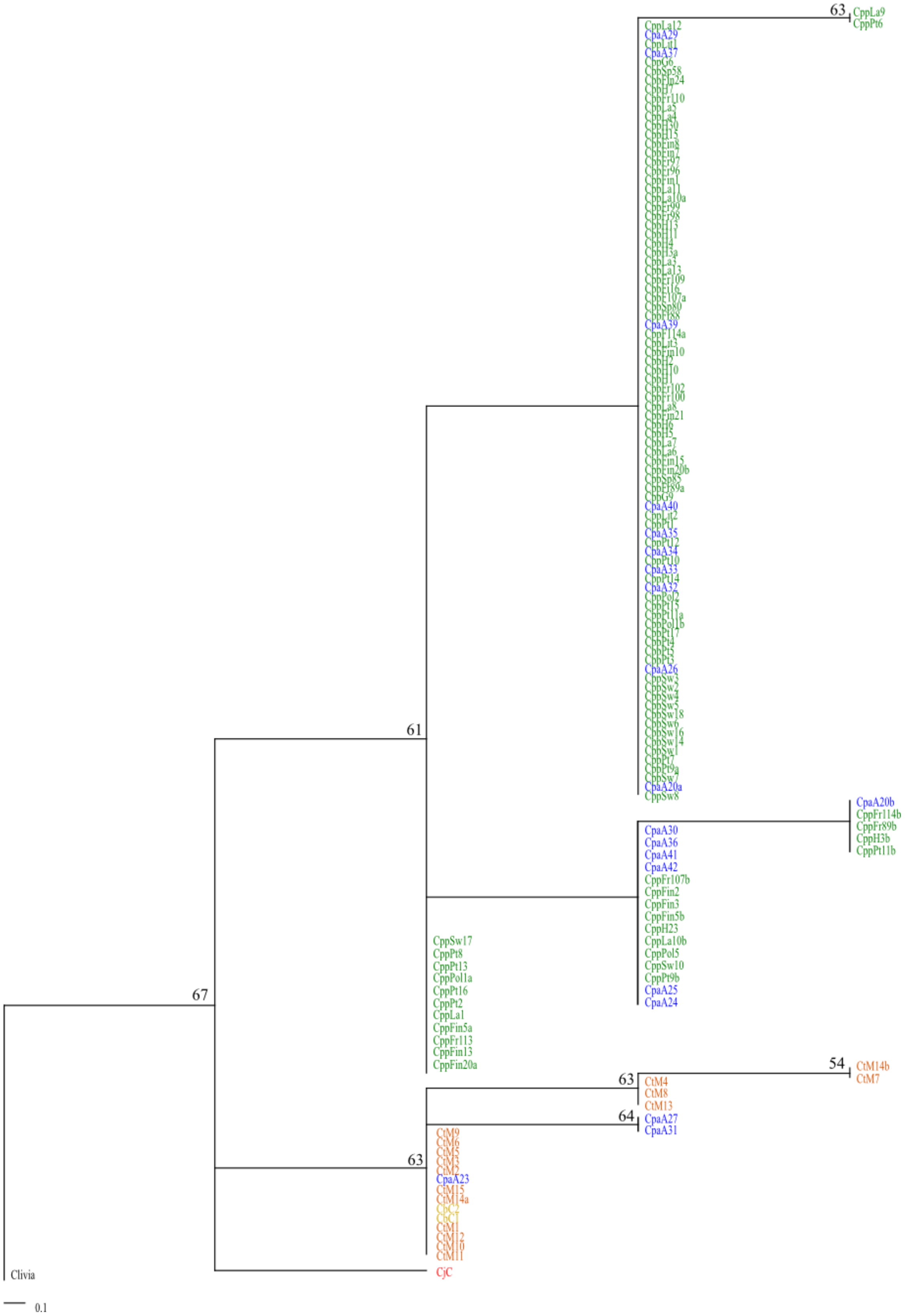

Figure 10: Maximum parsimony tree based on GHR intron 9 haplotypic data (one out of fifty-four trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refers to both haplotypes of a sample with ambiguous bases in this fragment.

The median-joining network (Figure 11) clearly shows the relationships among the individuals and their organization into haplotypes. The circle between four different haplotypes means the existence of parallel mutations, also known as homoplasies, among *C. p. palumbus* and *C. p. azorica*.

In Appendix VI, Figure 41 is possible to observe the median-joining network from GHR intron 9 data with heterozygous bases dealt as ambiguities.

From both cladograms and networks, a clear pattern emerges with an early differentiation of *C. junoniae*, later of *C. trocaz, C. bollii* and some *C. p. azorica* samples, and then *C. p. palumbus* and most of *C. p. azorica* samples.

From this fragment, it is likely that two colonization events may have occurred in the Azores: one from a pigeon near *C. trocaz* and *C. bollii* or the ancestor of both from Madeira archipelago or even the Canary Islands (less probable), and a more recent colonization from the mainland populations of *C. p. palumbus*, as it is alleged elsewhere.

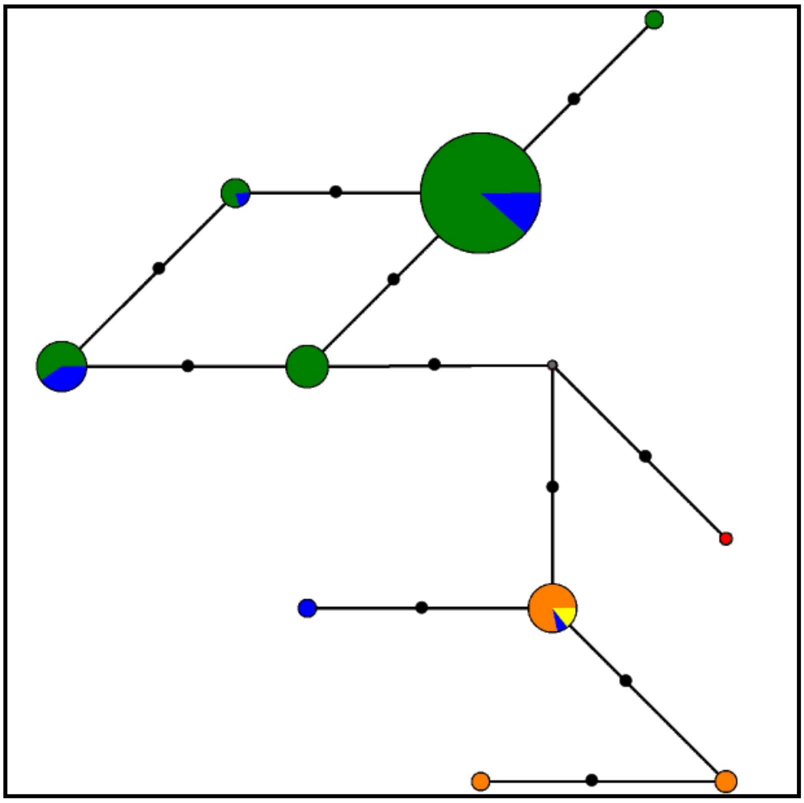

Figure 11: Median-joining network from GHR intron 9 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

4.3.2 RP40 Intron 5

Figure 12 shows the phylogenetic relationships between the studied species according to the RP40 intron 5. Unlike GHR intron 9, this fragment does not distinguish three main clades, but instead it makes a single polytomy with *C. p. palumbus*, *C. p. azorica* and *C. trocaz*, where *C. bollii* is also present. However, most *C. trocaz* samples contain two different haplotypes and come fairly well separated from the rest. The lack of resolution to at least distinguish species makes this intron not as reliable to build a phylogeny as the previous analyzed GHR intron 9.

In all respects, both the consensus tree (Appendix V, Figure 30) and the phylogram (Figure 12) are consistent.

The phased phylogram (Figure 13) is also coherent with the respective consensus tree (Appendix V, Figure 31). The phylogram maintains the main topology than the tree with ambiguities (Figure 12) since it also shows a large polytomy with samples from three taxa: *C. p. palumbus*, *C. p. azorica* and *C. trocaz*.

Comparing GHR and RP40, more ambiguous sites were found in the latter intron, and they were present in samples from all species, with the exception of *C. junoniae*.

Depicting a median-joining network, Figure 14 makes a bit more clear the relationships among the taxa, but the existence of homoplasies is likely on the basis of more than two different mutational paths connecting haplotypes in the network. Further, *C. p. palumbus*, *C. p. azorica* and *C. trocaz* share a haplotype, with a central position in the network.

In Appendix VI, Figure 42, where the median-joining network was built using heterozygous bases dealt as ambiguities, didn’t result in a much different graphic.

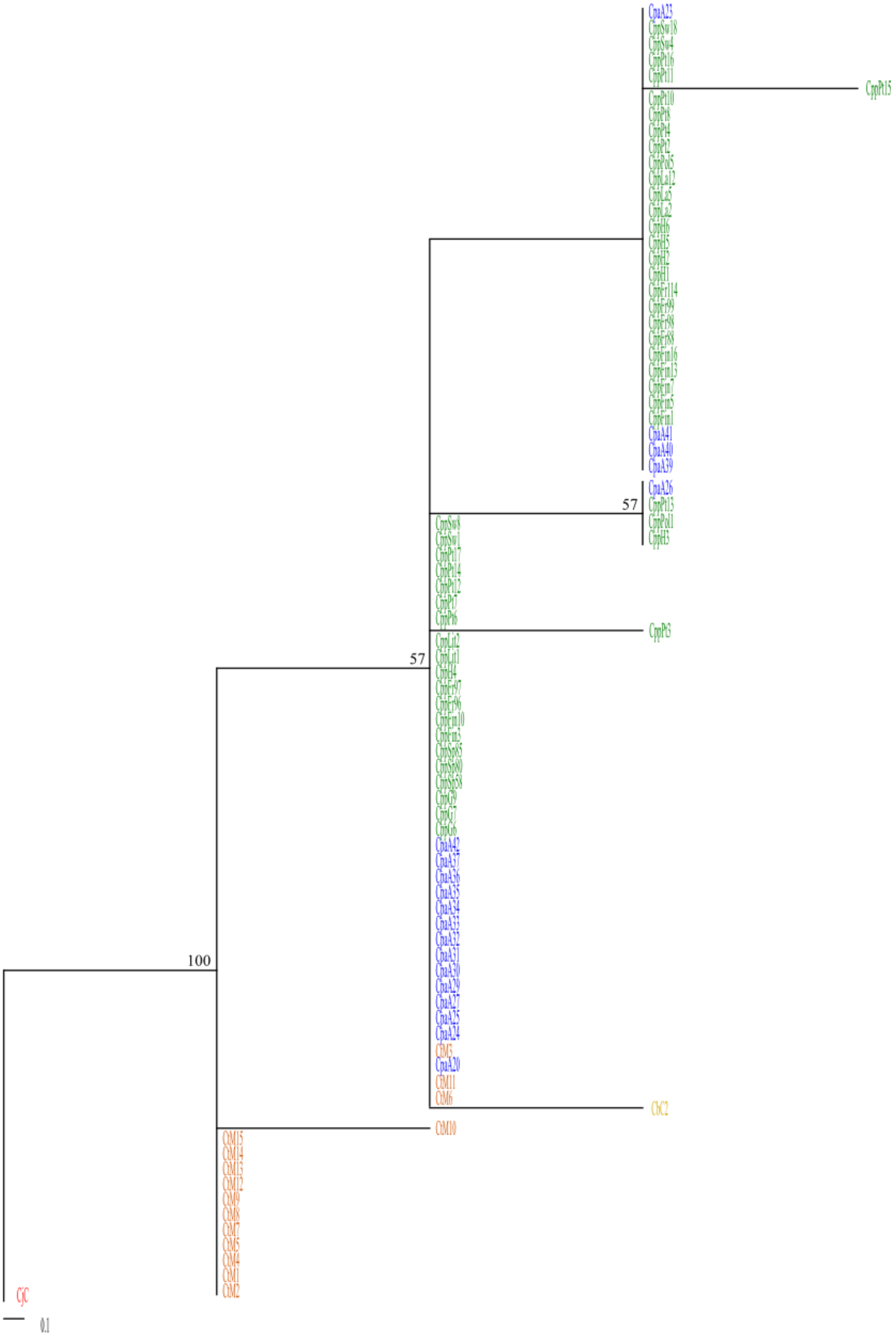

Figure 12: Maximum parsimony tree based on RP40 intron 5 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

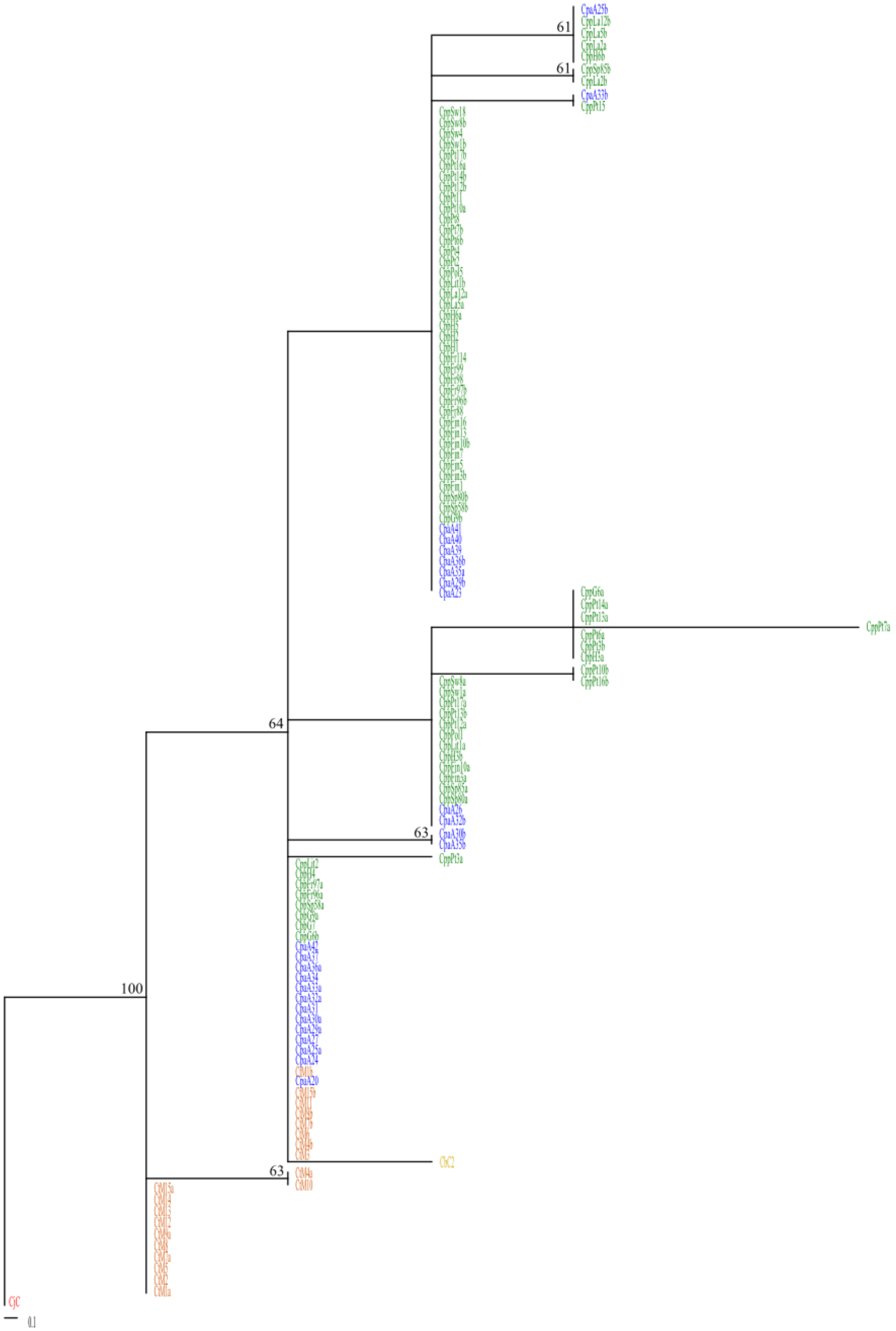

Figure 13: Maximum parsimony tree based on RP40 intron 5 haplotypic data (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refers to both haplotypes of a sample with ambiguous bases in this fragment.

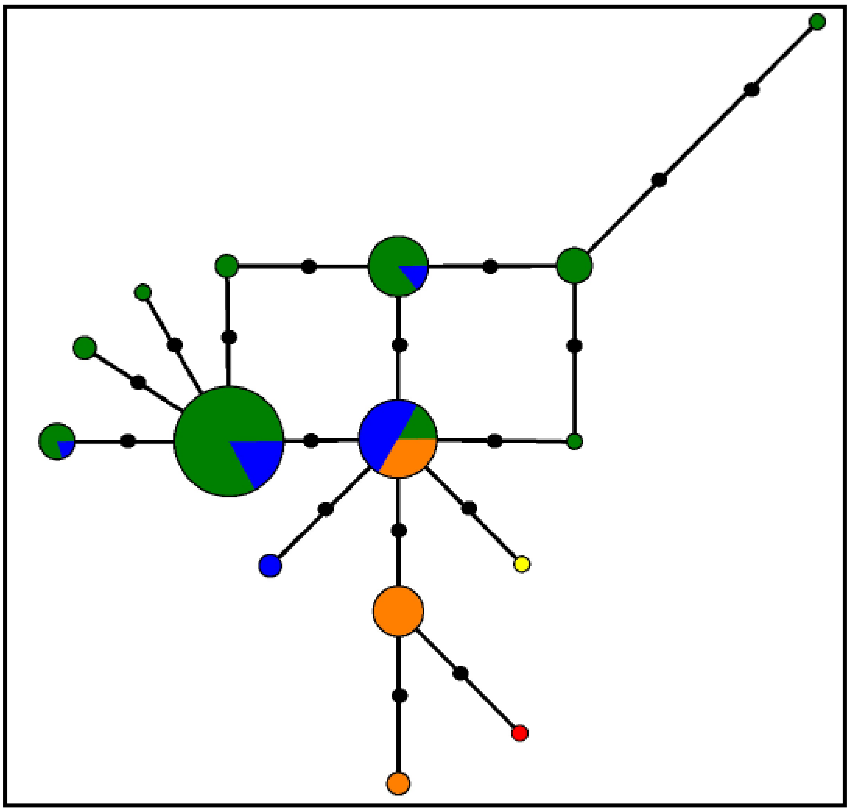

Figure 14: Median-joining network from RP40 intron 5 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. Black circles correspond to mutation events.

4.3.3 TROP Intron 5

The heterozygous parsimony tree based on TROP intron 5 (Figure 15) makes a clear distinction between *C. p. palumbus* and *C. trocaz. C. p. azorica* samples are shared by both groups. Two samples of *C. p. azorica* are phylogenetically closer to *C. p. palumbus*, whereas five cluster with *C. trocaz* and *C. bollii*. CpaA20 and CpaA31 come-up as a basal group.

Once again, *C. p. palumbus* samples show little differentiation, which is consistent with its migratory behavior.

With the exception of positions assumed by CpaA20 and CpaA31 in the consensus cladogram (Appendix V, Figure 32), the remaining topology is consistent with the displayed tree.

Both phased (Figure 16) and consensus tree (Appendix V, Figure 33) are coherent with exception of haplotypes CpaA31a, CtM4b, CtM14b and CtM15b, which appear as polytomic in the consensus cladogram.

Interesting results are revealed by the tree with haplotypic data (Figure 16). Observing each haplotype position from a single sample (represented by *a* and *b*), some *C. p. azorica* haplotypes appear in different groups – CpaA20a, CpaA20b, CpaA26a, CpaA26b, CpaA31a and CpaA31b. CpaA20 has two haplotypes; (*a*) appears in the *C. p. palumbus* clade, while (*b*) clusters with *C. trocaz* and *C. bollii*. The same occurs with CpaA26, while in Cpa31 the phenomenon is somewhat dissimilar, since one allele is in the *C. p. palumbus* clade, but the other forms a distinct group, also distinct from *C. trocaz* and *C. bollii*.

A likely explanation for this dichotomy would be a double origin for these haplotypes - Europe mainland and Madeira or Canary Islands. Due to a shorter distance from a possible source population, a hypothetical colonization from Madeira seems to have had more chances to occur than from the Canary Islands.

The median-joining network (Figure 17) elucidates the origin of *C. p. azorica* haplotypes. Homoplasies are still easily recognizable, but no haplotype is shared between *C. p. azorica* and *C. trocaz*.

Finally, Appendix VI, Figure 43 presents the median-joining network of TROP intron 5 data with heterozygous bases dealt as ambiguities.

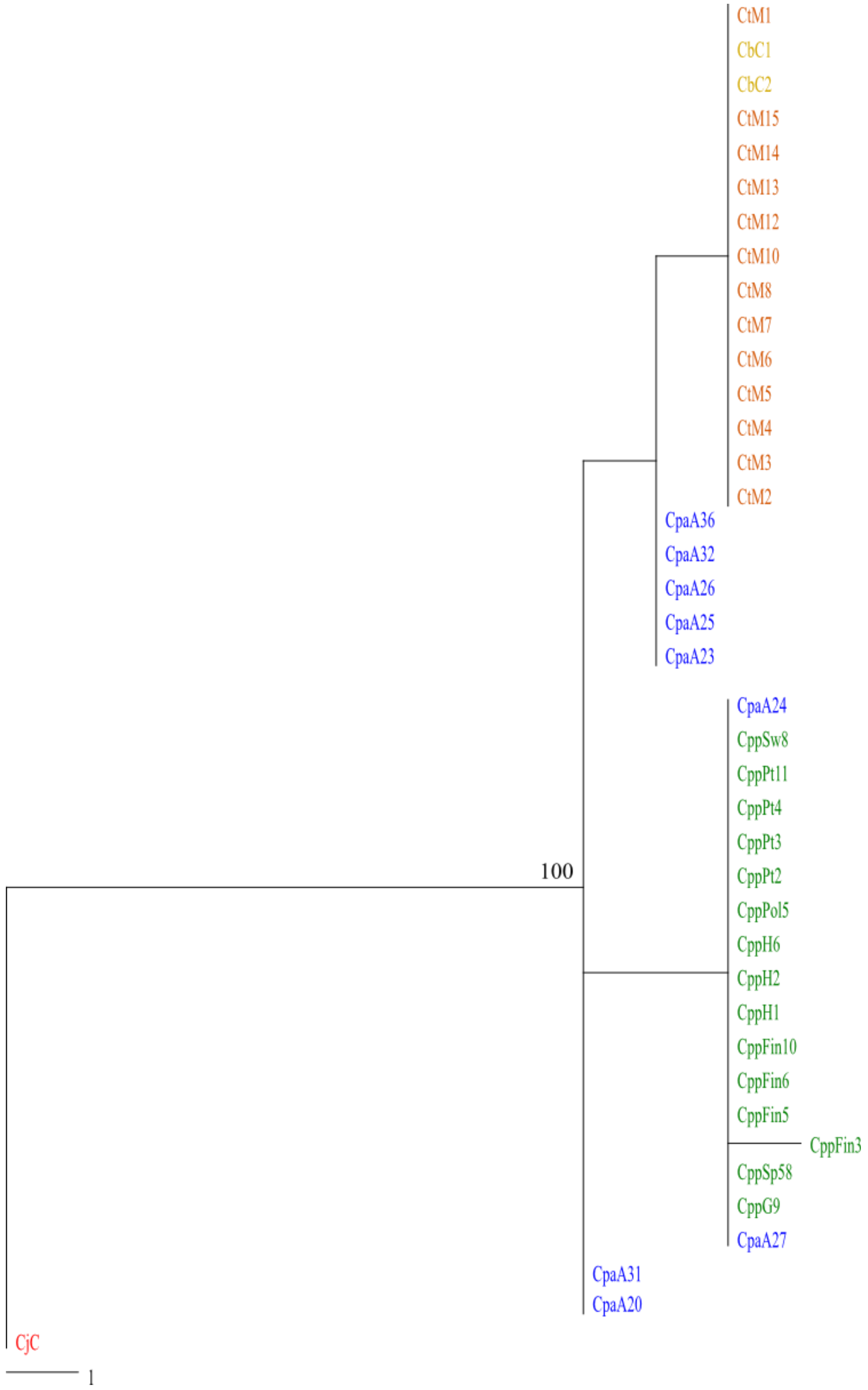

Figure 15: Maximum parsimony tree based on TROP intron 5 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

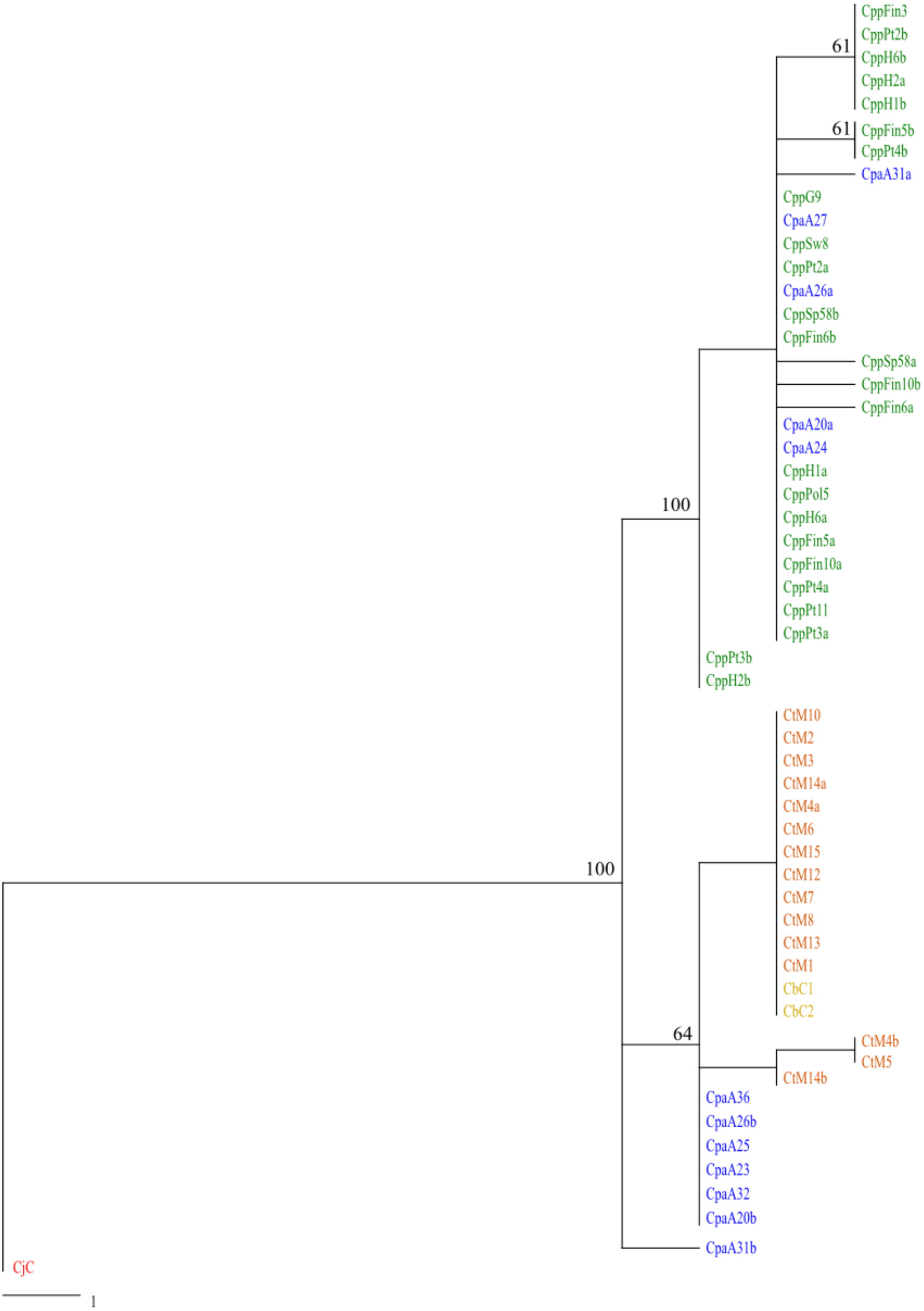

Figure 16: Maximum parsimony tree based on TROP intron 5 haplotypic data (one out of two hundred and seventy trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refers to both haplotypes of a sample with ambiguous bases in this fragment.

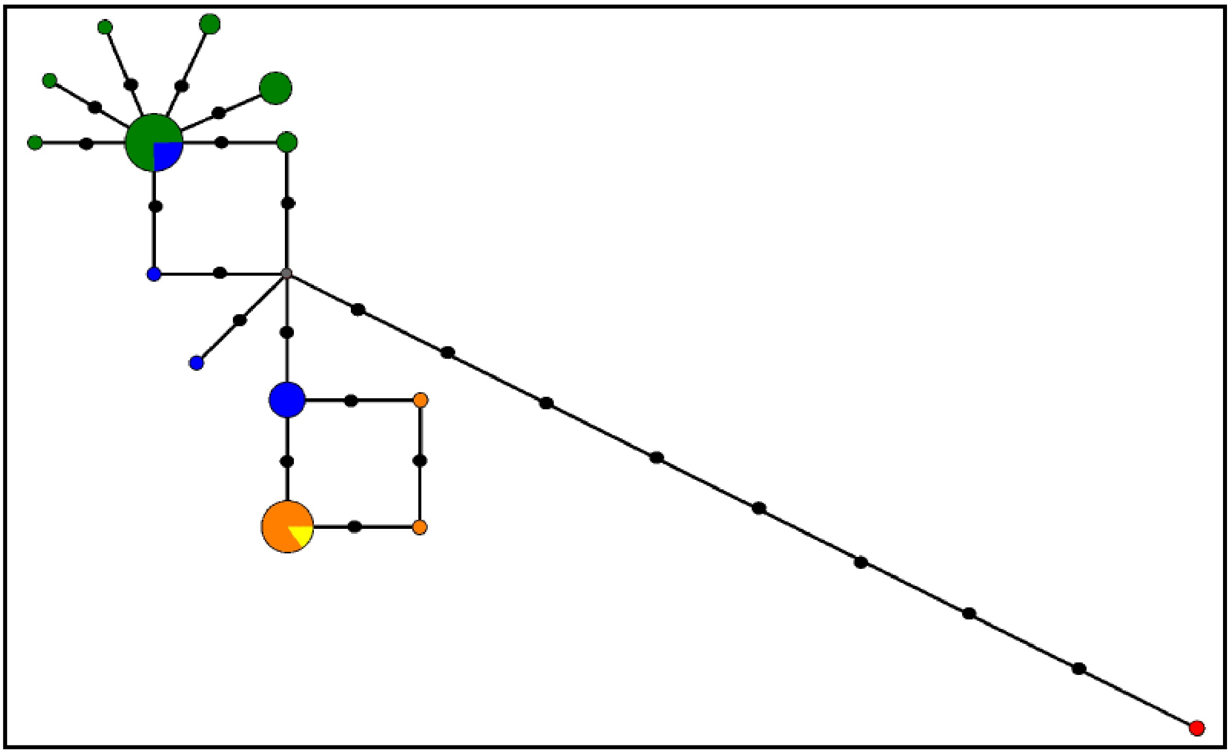

Figure 17: Median-joining network from TROP intron 5 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

4.3.4 β-FIB Intron 7

Due to the existence of samples with heterozygous length variation on β-FIB intron 7, two sequences with both length options appear in the tree, although they only differ in the presence or absence of the indel. Among the *C. trocaz* samples, four out of thirteen show length variation; among the eleven samples of *C. p. azorica*, four samples also show this polymorphism; within the fifteen samples of *C. p. palumbus*, four are heterozygous for length variation.

In the phylogram with heterozygous bases dealt as ambiguities (Figure 18), three groups emerge: the basal group is formed by *C. junoniae*; a second includes samples of *C. trocaz*, *C. bollii* and also *C. p. azorica*; finally, samples of *C. p. palumbus* and *C. p. azorica* constitutes the last group.

Three azorean specimens showing length variation appear in distinct clades (CpaA20lvh1, CpaA20lvh2, CpaA25lvh1, CpaA25lvh2, CpaA39lvh1 and CpaA39lvh2), according to the presence or absence of the indel. On the other hand, CpaA29 haplotypes (CpaA29lvh1 and CpaA29lvh2) are in the same clade.

In the consensus tree (Appendix V, Figure 34) the general topology is retained but a few differences may be taken into account. For instance, CpaA20, CpaA25, CpaA26, CpaA36, CpaA39 and CppPt11lvh2 are not polytomic. This has no major consequences for phylogenetic inference.

Confronting the phased phylogram (Figure 19) and the respective consensus cladogram (Appendix V, Figure 35), some differences can be found in the group composed by samples of *C. p. palumbus* and *C. p. azorica*, varying the position of some samples.

The phased phylogram is however quite interesting in that it provides novel insights on the phylogenetic relationships among atlantic pigeons. For instance, the fact that samples CpaA20, CpaA25 and CpaA39 (each one represented by four haplotypes in the tree) appear separated (essentially due to their base composition, since all the *b* haplotypes are among the *C. trocaz* and *C. bollii* samples, while the *a* haplotypes clusters with woodpigeon from Europe mainland) contradicts the pre-conceived idea of a single pigeon colonization of the Azores by *C. p. palumbus*. This suggests once more a double colonization of these islands both from the European mainland and from Madeira or Canary Islands.

In the median-joining network (Figure 20), the existence of several homoplasies and hypothetical haplotypes might have contributed to the network complexity as well as the presence of several median vectors.

The median-joining network from β-FIB intron 7 data with heterozygous bases dealt as ambiguities is presented in Appendix VI, Figure 44.

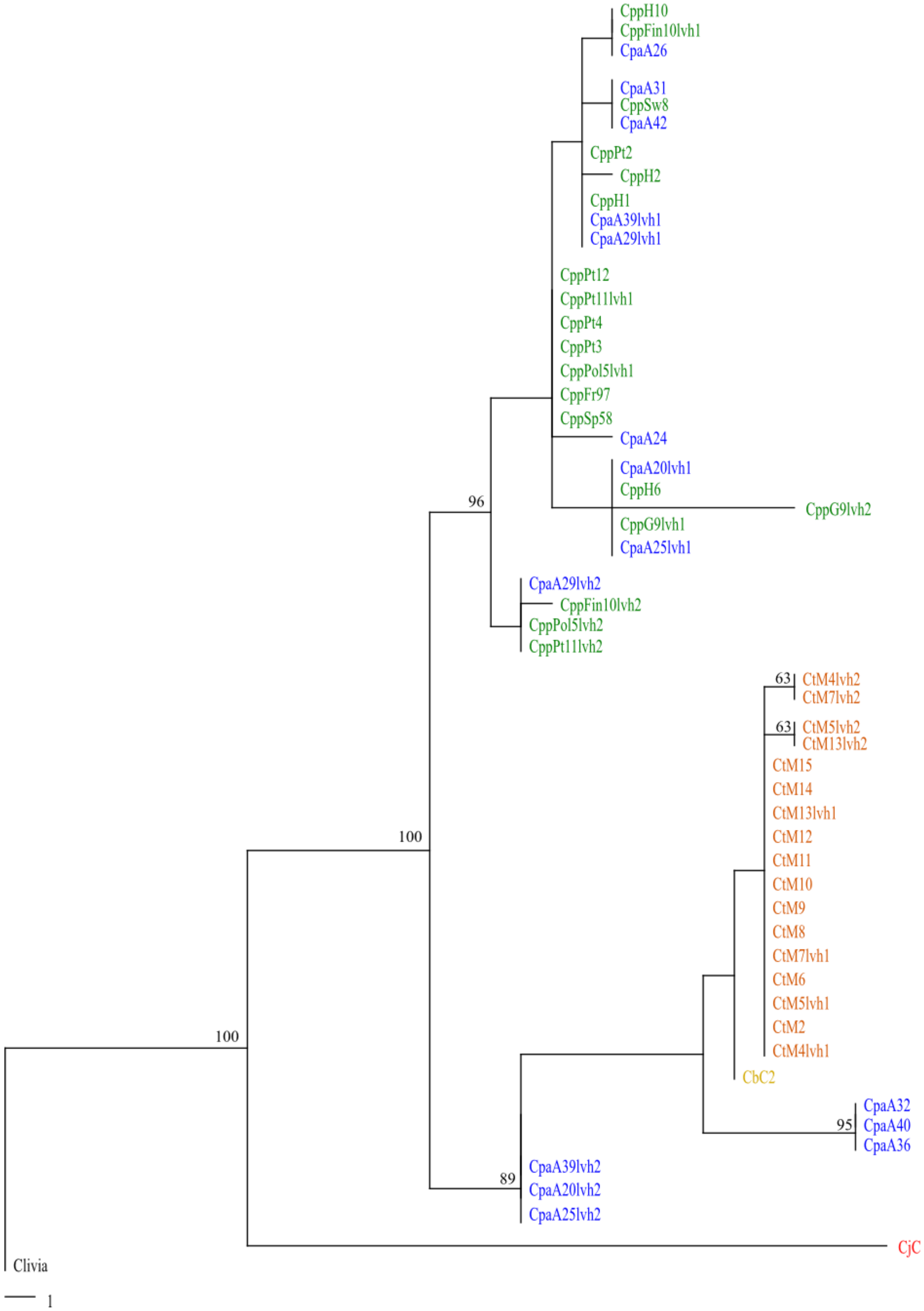

Figure 18: Maximum parsimony tree based on β-FIB intron 7 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *lvh1* and *lvh2* refer to the both possible haplotypes when the sample has a length variation in the fragment.

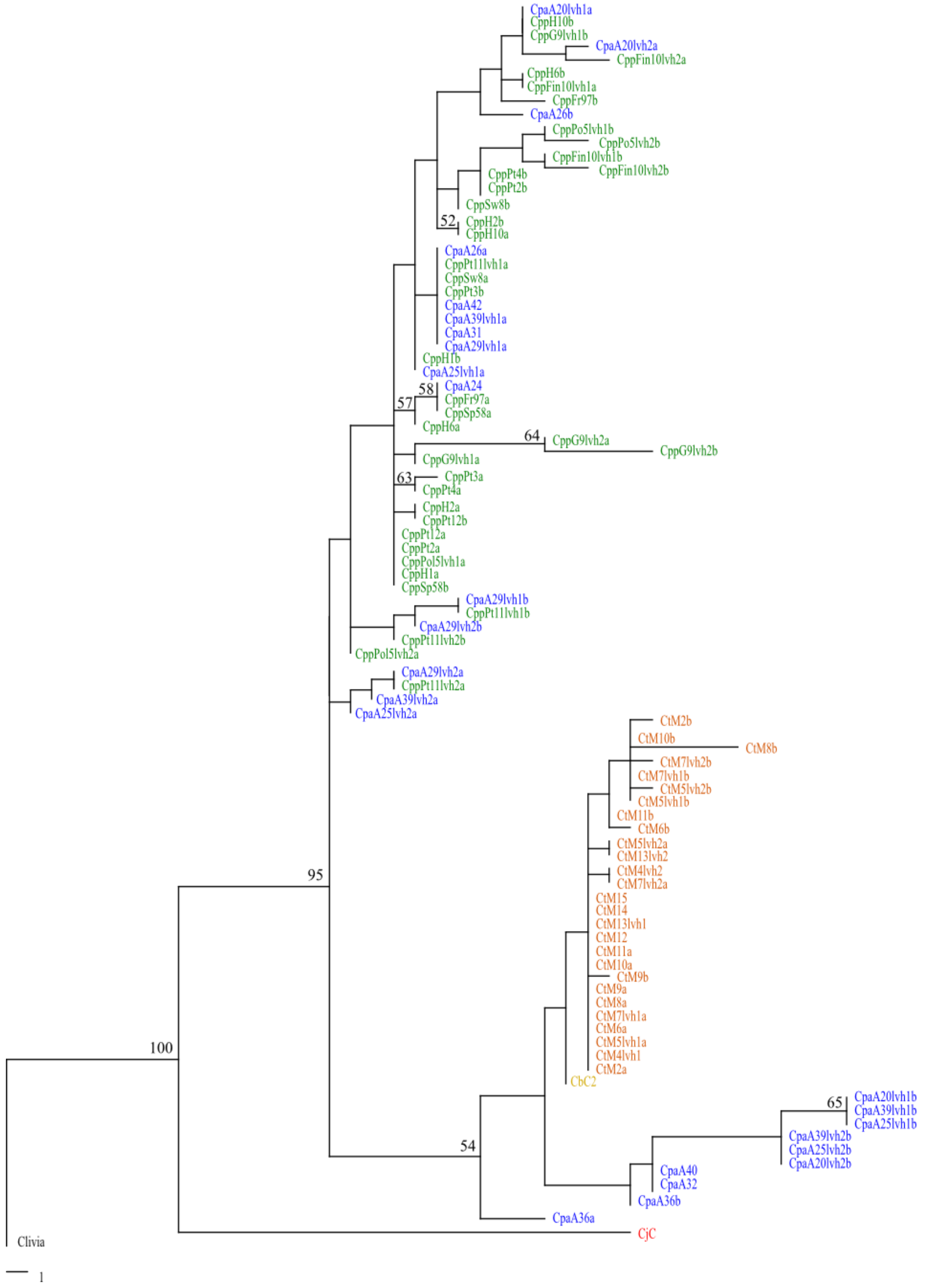

Figure 19: Maximum parsimony tree based on β-FIB intron 7 haplotypic data (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. The samples with both heterozygous sites and length variation, four possible options are shown - *lvh1a*, *lvh1b*, *lvh2a and lvh2b*.

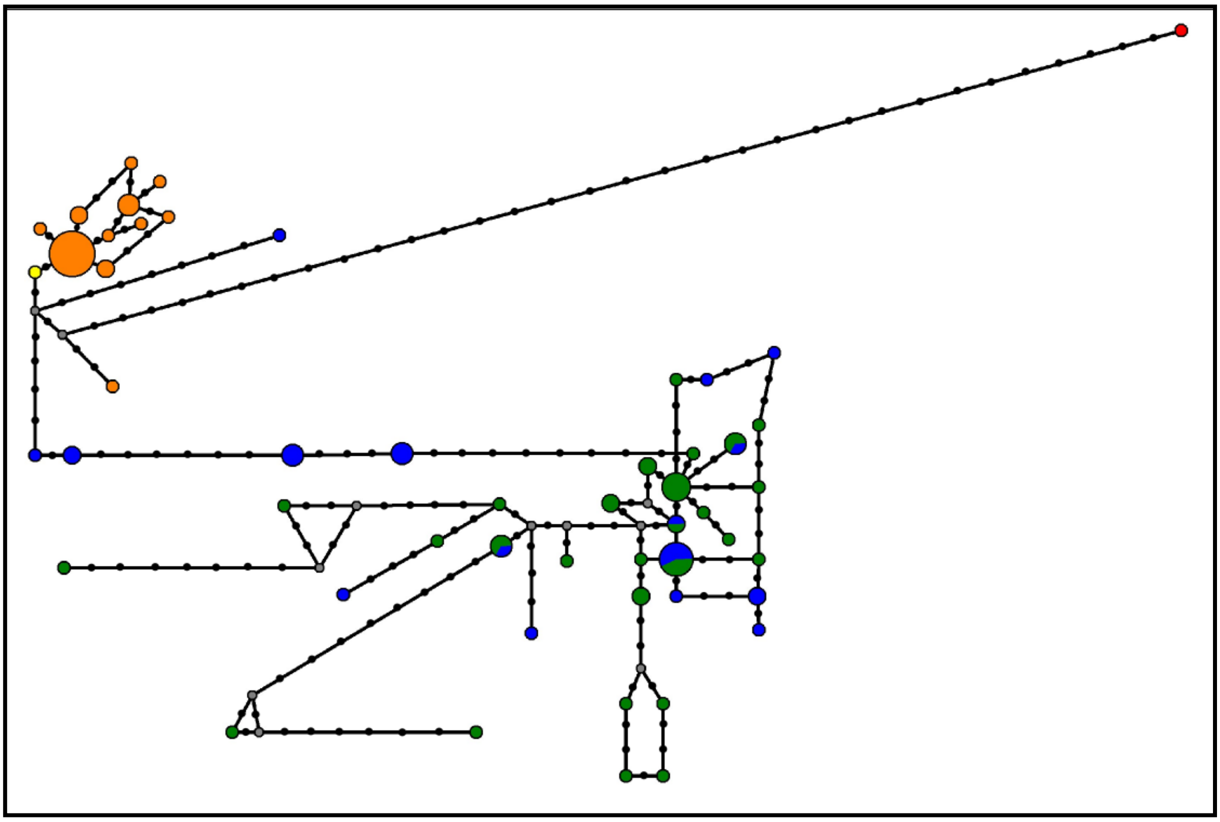

Figure 20: Median-joining network from β-FIB intron 7 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. Median vectors are represented by grey circles. Black circles correspond to mutation events.

4.3.5 TGF-β2 Intron 5

Both the consensus tree (Appendix V, figure 36) and the phylogram (Figure 21) of TGF-β2 intron 5 are very similar in making a clear distinction between *C. trocaz* and both *C. p. palumbus* and *C. p. azorica*, being the positions occupied by CtM1, CtM12, CtM13 and CtM14 the only discrepancies between them. Nonetheless, these little incongruences are not phylogenetically significant.

It is easy to verify in the phylogram regarding the haplotypic data (Figure 22) that almost all samples have ambiguous bases.

One haplotype from sample CpaA29 (CpaA29b) appears in the *C. trocaz* clade, what is not verified in the tree with the heterozygous sequences. In spite of the low bootstrap

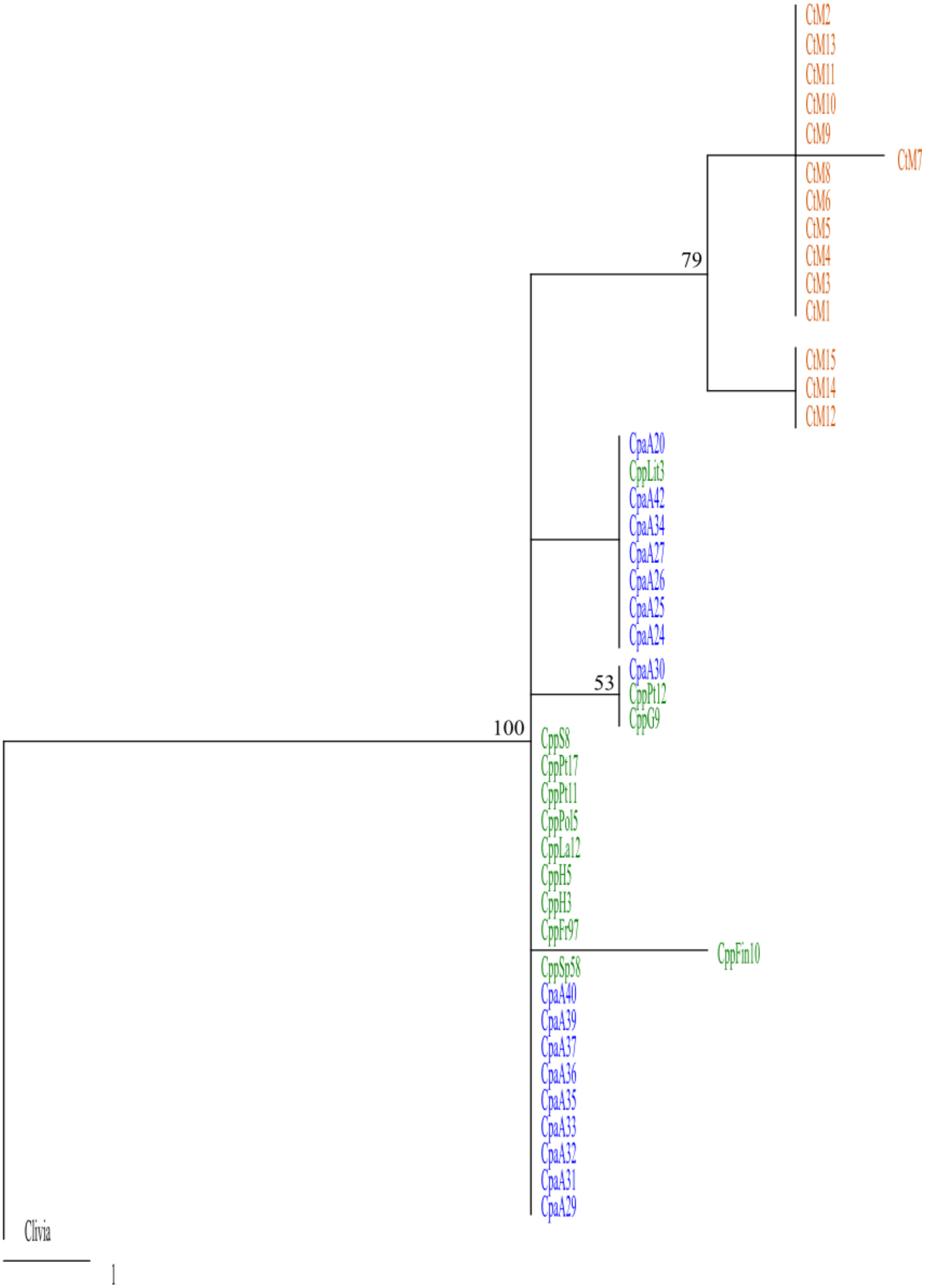

Figure 21: Maximum parsimony tree based on TGF-β2 intron 5 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*.

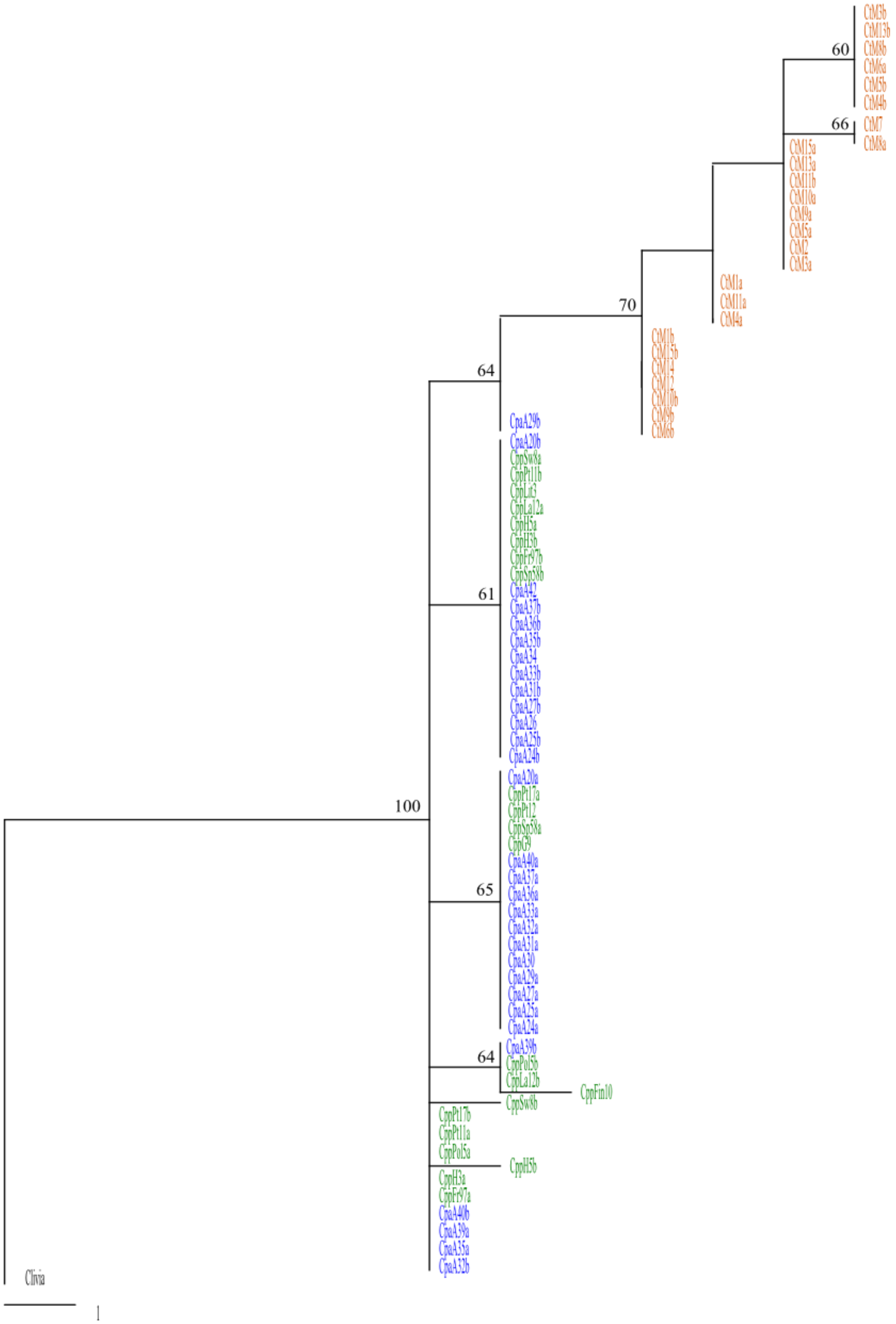

Figure 22: Maximum parsimony tree based on TGF-β2 intron 5 haplotypic data (one out of thirty trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

support value, this occurs in all obtained trees, and it is confirmed in the consensus cladogram (Appendix V, Figure 37).

Both the median-joining network (Figure 23) and the phylogram (Figure 22) exhibit a similar result, distinguishing *C. trocaz* and both *C. p. palumbus* and *C. p. azorica*.

In Appendix VI, Figure 45 shows the median-joining network from TGF-β2 intron 5 data with heterozygous bases dealt as ambiguities.

Data obtained with this nuclear marker shows some discrepancies from the previously analyzed, since all *C. p. azorica* samples cluster with *C. p. palumbus* (with the exception of CpaA29b in the haplotypic phylogram). This may be due to a complete lineage sorting.

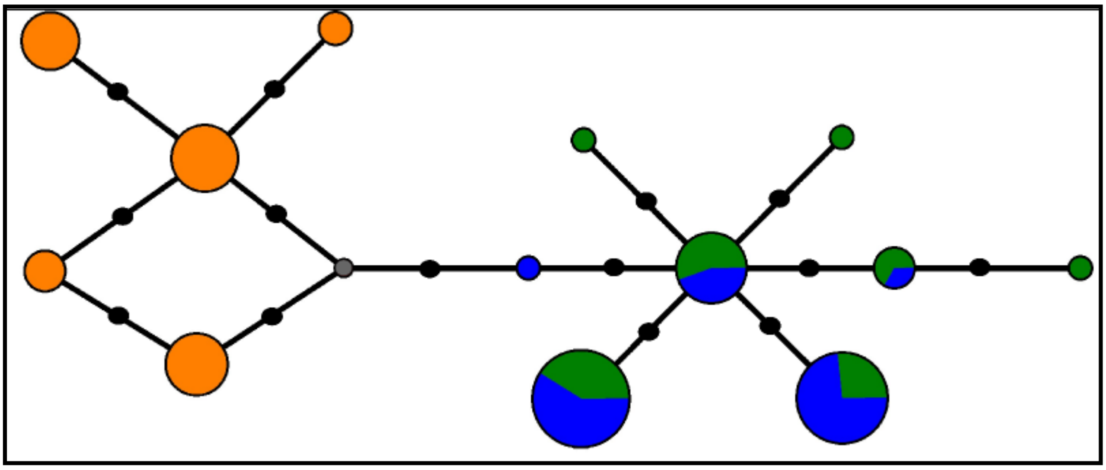

Figure 23: Median-joining network from TGF-β2 intron 5 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

4.3.6 IRF2 Intron 2

Both phylograms (Figures 24 and 25) and respective consensus trees (Appendix V, Figures 38 and 39) obtained from IRF2 intron 2 data (unphased and phased, respectively) are very similar in not even allowing to differentiate *C. p. palumbus* from *C. trocaz,* which are in turn mixed with *C. p. azorica*.

Figure 24: Maximum parsimony tree based on IRF2 intron 2 sequence data with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*.

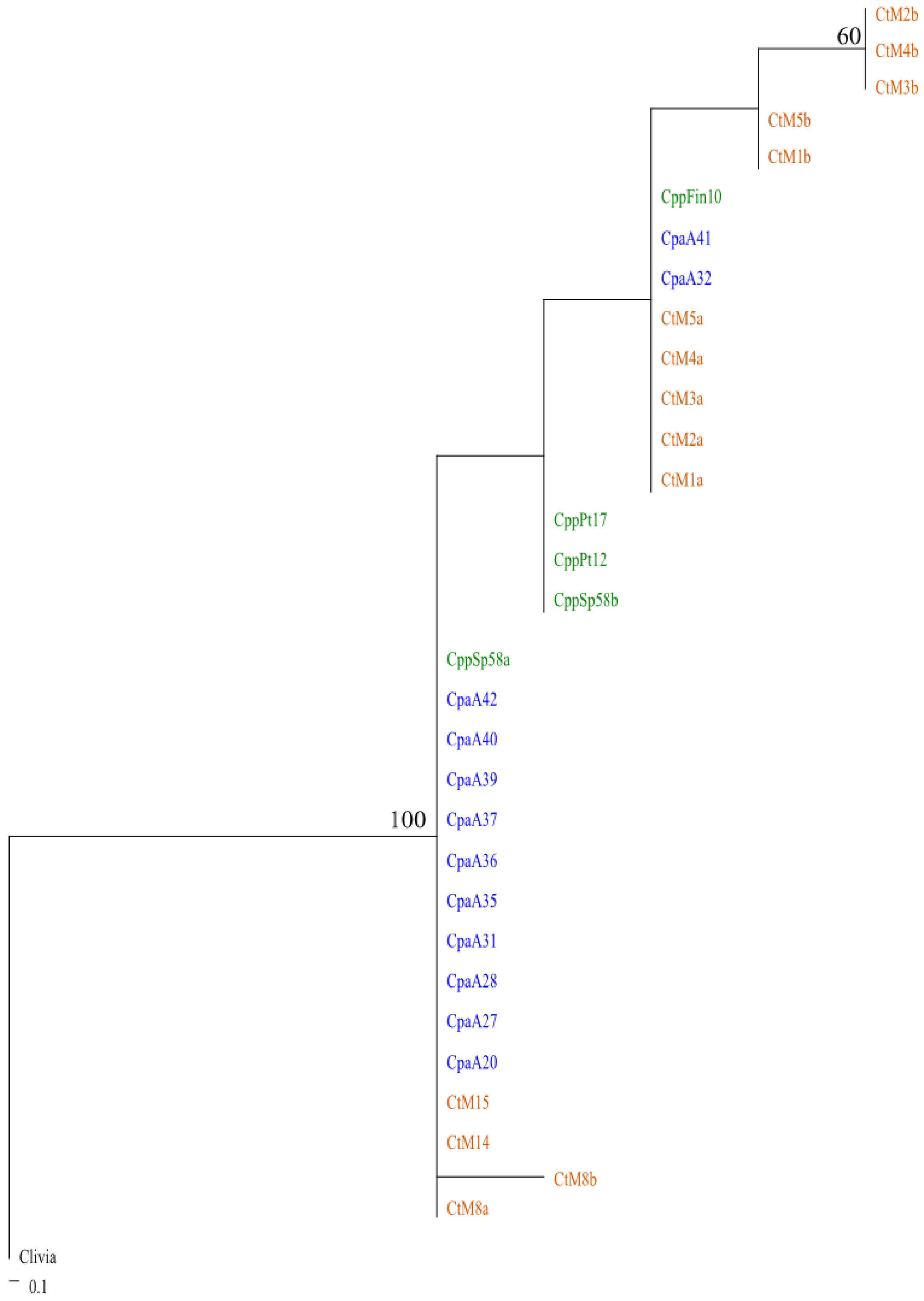

Figure 25: Maximum parsimony tree based on IRF2 intron 2 haplotypic data (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

Several samples from this fragment are shorter than desirable. So, the data matrix had several ambiguities (unknown bases, coded as “N”). Since some programs do not deal appropriately with this, those bases were removed from the data matrix. In the process, some differences among the sequences were not included, and the number of haplotypes was consequently reduced: only three haplotypes could be retrieved in the median-joining network (Figure 26). Here, one haplotype is common to *C. p. palumbus*, *C p. azorica* and some *C. trocaz* samples, while the other two are exclusive to *C. trocaz*.

Regarding the network with heterozygous bases dealt as ambiguities, and according to the previous explanation, the final data matrix was shorter and has less than three different haplotypes, preventing the network inference.

This intron was not considered further in the concatenated dataset.

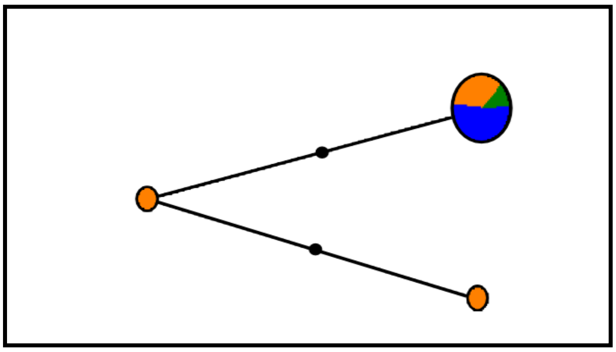

Figure 26: Median-joining network from IRF2 intron 2 haplotypic data. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. Black circles correspond to mutation events.

4.3.7 Introns Data – Further Considerations

Observing each phylogram independently, it is evident that a particular *C. p. azorica* sample is not always in the same clade or group, according to different molecular markers used.

When analyzing phylograms with heterozygous positions dealt as ambiguities, CpaA23 appears in the *C. trocaz* clade when considering fragment GHR, but it is in the *C. p. palumbus* clade of TROP phylogram. Note the position assumed by samples CpaA24 in TROP and β-FIB phylograms; CpaA31 in GHR, TROP and β-FIB phylograms; CpaA32 in

GHR, TROP and in β-FIB phylograms; CpaA36 in GHR, TROP and β-FIB phylograms, and also CpaA40 in GHR and β-FIB phylograms. However, some samples remain in the same position (CpaA26 in GHR, TROP and β-FIB phylograms, and CpaA29 in GHR and β-FIB phylograms).

These data suggest that one specimen may have genetic information with distinct origins (*C. p. palumbus* and *C. trocaz* or *C. bollii*), and this may be the case of some *C. p. azorica* samples.

4.3.8 Concatenated Data

The resulting tree from the concatenated dataset is shown on Figure 27. Some differences are found between this and the consensus tree (Appendix V, Figure 40), like different positions of samples CtM2, CtM5-1, CtM9 and CpaA39 (which is closer *C. trocaz* in the consensus cladogram). The other branches are congruent.

Once more two groups appear, being one formed by all *C. trocaz* samples and some *C. p. azorica* samples, and the other constituted by *C. p. palumbus* and several *C. p. azorica* samples. Two hypothetic haplotypes (CpaA25-2 and CpaA39-2) do not belong to any of the previous clades, and have a basal position to the remain samples.

These results are congruent with the ones obtained by some fragments solely (GHR, TROP and β-FIB) with *C. p. azorica* in the two main clades, suggesting a double origin of this insular woodpigeon population.

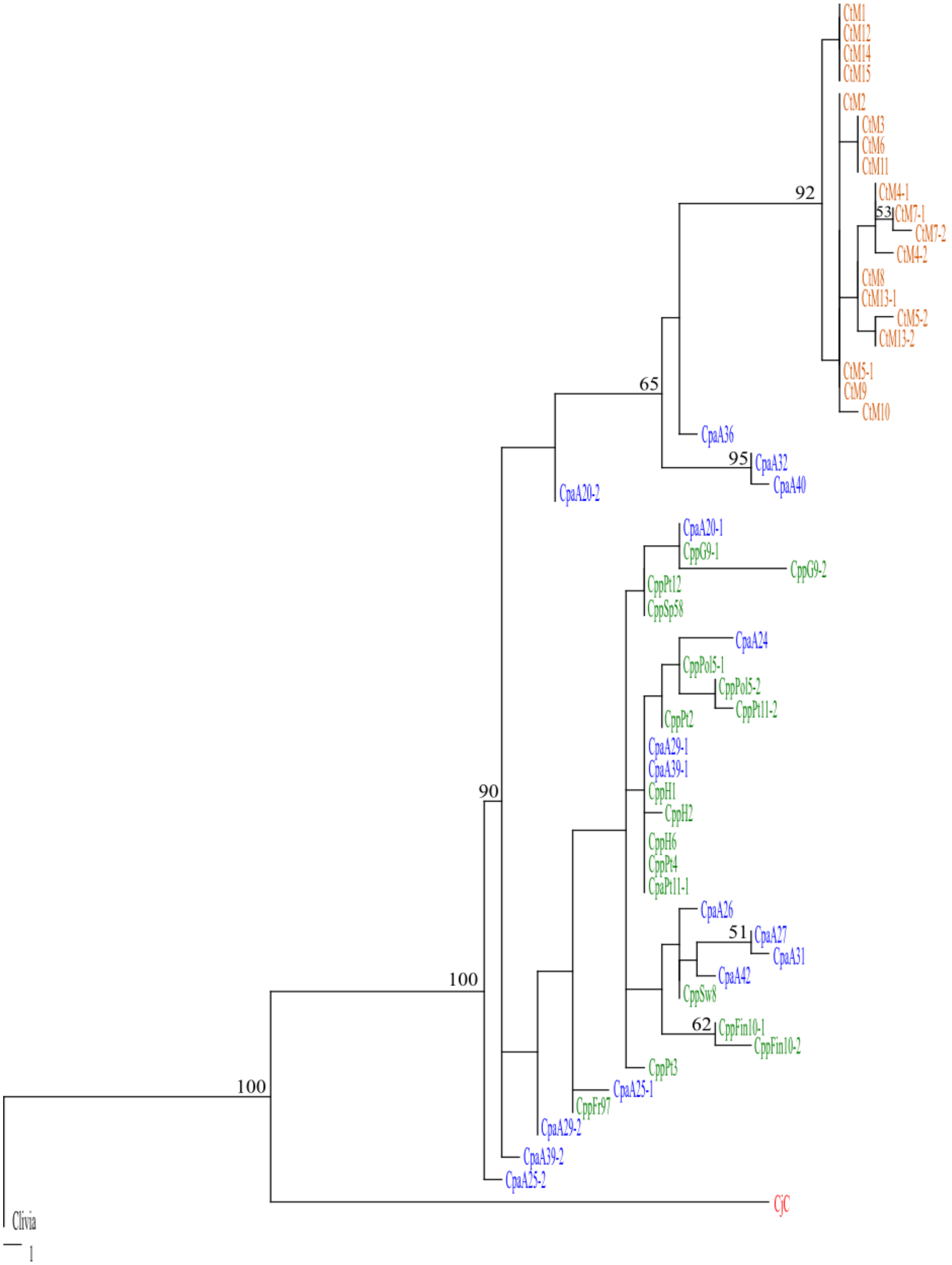

Figure 27: Maximum parsimony tree based on 5 fragments concatenated (GHR, RP40, TROP, β-FIB and TGF-β2), with heterozygous bases dealt as ambiguities (one out of one thousand trees with equal tree length). Branch support is given by MP bootstrap values (only values above 50% are shown). Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz* and red – *C. junoniae*. Numbers *1* and *2* refer to both concatenation options due to the presence of a length variation in β-FIB intron 7.

5. Final Remarks

From the phylogenetic and phylogeographic analyses undertaken in this work, considering the data from isolated genes and concatenated datasets either in the trees or networks, we can take some considerations on the natural history and evolution of pigeons in atlantic islands and nearby continental landmasses.

We can observe that gene trees are different from each other, and consequently gene trees and species trees are not the same, as argued by Nichols (2001). However, three out of six fragments used in this work show the same general pattern, as well as the concatenated phylogram.

The Laurel Pigeon, *C. junoniae*, had possibly arrived Canary Islands much earlier than *C. bollii* (Gonzalez *et. al.* 2009), within a timeframe between 30 and 5 million years before present. This hypothesis is extensively corroborated by most markers in this study, as it comes as an outgroup of the other species.

C. bollii (Canary Islands) and *C. trocaz* (Madeira) are phylogenetically closer, and they are probably the result of a simultaneous colonization event from a mainland population. In a recent study, it is argued that *C. bollii* arrived Canary Islands about 1.6 – 5 Myr ago (Gonzalez *et. al.* 2009). Due to molecular similarities shared between these two endemic atlantic pigeons, it is possible that the common ancestor of both species arrived at Madeira archipelago and Canary Islands about the same time, and that is the reason why they are phylogenetically related.

C. p. palumbus positions in phylogenetic trees and in median-joining networks show that there is no evidence of structure among samples from several geographic locations. This may be related with the migratory behavior of this species (Bea *et al.* 2003), which may be maintaining a panmictic population.

Our results show that *C. p. azorica* is not a monophyletic taxa, since its samples generally comes-up in two clades (one consisting of *C. p. palumbus* and other by *C. trocaz* and *C. bollii*). This had been previously demonstrated by Silva (2007), but is now confirmed on the basis of a much robust dataset (with more introns and samples).

Mitochondrial data suggest the inexistence of differentiation between *C. p. palumbus* and *C. p. azorica* (Grosso 2002; Duarte 2006). Thus, from this study we can notice that nuclear and mitochondrial markers have different phylogenetic signal.

In the insular woodpigeon population only prevail mtDNA from *C. p. palumbus*, while nDNA is not only similar to the European mainland population but also similar to *C. trocaz* and *C. bollii*, according to different samples.

The discrepancy between mtDNA and nDNA data could be due to female-biased dispersal, i.e., the imbalance in the dispersal behavior of females, leading to a higher dispersal of that particular sex. Thus, due to sex-biased dispersal different DNA regions are subject to different levels of gene flow as a consequence of their mode of inheritance – biparental, maternal or parental. It is known that several bird species have female-biased dispersal (Petit & Excoffier 2009), and it may be the case of genus *Columba*.

It is usually assumed that an insular species results from a single colonization event either from a nearby continent or from other islands (Emerson 2002). Nevertheless, this may not be the case in all situations.

Within the Azores archipelago, the contemporaneous population of woodpigeons seems to be the result of a double colonization. Azores was probably colonized first by either an ancestor of *C. trocaz* and *C. bollii* or one of them, diverging from both due to geographic isolation. Nuclear data indicates that there was a second colonization, this time from woodpigeon mainland population, with subsequent admixture.

Considering island age and geographical distance between each archipelago and Europe mainland (see Figure 1), this hypothesis of a double colonization and admixture seems solid. Unlike azorean population, *C. trocaz* and *C. bollii* populations have remained isolated from Europe mainland woodpigeon, maybe due to their southern position further away from the core of migration route of northern European woodpigeons populations. With an admixture pattern it is comprehensible why *C. trocaz*, *C. bollii* and some *C. p. azorica* are phylogenetically similar, while other *C. p. azorica* are seem to be closer to *C. p. palumbus*.

Hybridization can be the explanation for differences among mtDNA and nDNA: hybridization may have conducted to introgression of mtDNA from *C. p. palumbus* into the initial azorean population, where the newly arrived *C. p. palumbus* mtDNA form prevailed.

Gene transfer among species during hybridization is widely accepted in plants, but it is considered a relatively unusual phenomenon in animals (Zakharov *et. al.* 2009). Some interesting events of introgression are documented, such as the case of genus *Hyles* (Lepidoptera: Sphingidae) (Hundsdoerfer *et al.* 2009) and Darwin's finches (Sato *et al.* 1999).

We may then conclude that instead of a mere isolated woodpigeon population, *C. p. azorica* represents a much more complex and fascinating event in pigeon evolution than it was previously believed.

6. References

Abrantes PCS (2000) *Caracterização da estrutura genética de quarto populações de* Columba palumbus*, através da análise de microssatélites*. Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Agudo R, Rico C, Hiraldo F, Donázar JA (2011) Evidence of connectivity between continental and differentiated insular populations in a highly mobile species. *Diversity and Distributions*, **17**, 1-12.

Arechavaleta M, Zurita N, Marrero MC, Martin JL (2005) *Lista preliminary de species silvestres de Cabo Verde (hongos, plantas y animals terrestres)*. Consejeria de Medio Ambiente y Ordenación Terretorial, Gobierno de Canarias.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3-15.

Bankovics A (2001) The migration of Wood Pigeon (*Columba palumbus*) and Turtle Dove (*Streptopelia turtur*) in Hungary. *Naturzale*, **16**, 83-93.

Bandelt H-J, Forster P, Röhl A (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

BBC. www.bbc.co.uk/nature/wildfacts/factfiles/255.shtml. (18th March 2011).

Bea A, Beitia R, Fernández JM (2003) The census and distribution of wintering woodpigeons *Columba palumbus* in the Iberian peninsula. *Ornis Hungarica*, **12-13**, 157-167.

Bellemain E, Bermingham E, Ricklefs RE (2008) The dynamic evolutionary history of the bananaquit (*Coereba flaveola*) in the Caribbean revealed by a multigene analysis. *BMC Evolutionary Biology*, **8**, 1-14.

Bollmer JL, Kimball RT, Whiteman NK, Sarasola JH, Parker PG (2006) Phylogeography of the Galápagos hawk (*Buteo galapagoensis*): A recent arrival to the Galápagos Islands. *Molecular Phylogenetics and Evolution*, **39**, 237-247.

Bonilla AJ, Braun EL, Kimball RT (2010) Comparative molecular evolution and phylogenetic utility of 3'-UTRs and introns in Galliformes. *Molecular Phylogenetics and Evolution*, **56**, 536-542.

Borges PAV, Amorim IR, Cunha R, Gabriel R, Martins AF, Silva L, Costa A, Vieira V (2009) Azores – Biology. In Gillespie & Clague (Eds.), *Encyclopedia of Islands*, pp 71, 2009, University of California Press, California.

Brandström M, Ellegren H (2007) The Genomic Landscape of Short Insertion and Deletion Polymorphisms in the Chicken (*Gallus gallus*) Genome: A High Frequency of Deletions in Tandem Duplicates. *Genetics*, **176**, 1691-1701.

Brito PH, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439-455.

Brookes AJ (1999) The essence of SNPs. *Gene*, **234**, 177-186.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *TRENDS in Ecology and Evolution*, **18**, 249-256.

Brumfield RT, Liu L, Lum DE, Edwards SV (2008) Comparison of Species Tree Methods for Reconstructing the Phylogeny of Bearded Manakins (Aves: Pipridae, *Manacus*) from Multilocus Sequence Data. *Systematic Biology,* **57**, 719-731.

Bruun B, Delin H, Svenson L (1993) *Aves de Portugal e Europa*. pp 166-167, FAPAS, Porto.

Cabral MJ, Almeida J, Almeida PR, Dellinger T, Almeida NF, Oliveira ME, Palmeirim J, Queiroz AI, Rogado L, Reis MS (2005) *Livro Vermelho dos Vertebrados de Portugal*. pp 225-232, Instituto de Conservação da Natureza, Lisboa, Portugal.

Clark DA, Coker R (1998) Transforming growth factor-beta (TGF-β). *The International Journal of Biochemistry & Cell Biology*, **30**, 293-298.

Correia-Fagundes C, Romano H (2011) Trocaz Pigeon Columba trocaz. Available at www.madeirabirds.com/trocaz_pigeon_columba_trocaz. (Accessed: 15th March 2011).

Creer S, Pook CE, Malhotra A, Thorpe RS (2006) Optimal Intron Analyses in the *Trimeresurus* Radiation of Asian Pitvipers. *Systematic Biology*, **55**, 57-72.

Creer S (2007) Choosing and Using Introns in Molecular Phylogenetics. *Evolutionary Bioinformatics*, **3**, 99-108.

Dietzen C, Garcia-del-Rey E, Castro GD, Wink M (2008) Phylogenetic differentiation of *Sylvia* species (Aves: Passeriformes) of the Atlantic islands (Macaronesia) based on mitochondrial DNA sequence data and morphometrics. *Biological Journal of the Linnean Society*, **95**, 157-174.

Donázar JA, Gangoso L, Forero MG, Juste J (2005) Presence, richness and extinction of birds of prey in the Mediterranean and Macaronesian islands. *Journal of Biogeography*, **32**, 1701-1713.

Duarte MASCT (2006) *Filogenia e Genética Populacional dos pombos europeus do género* Columba *Linnaeus, 1758 recorrendo a diferentes marcadores moleculares.* Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Elias G, Reino L, Silva T, Geraldes P, Tomé R (1998) *Atlas das Aves Invernantes do Baixo Alentejo*. SPEA, Lisboa.

Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecolo*gy, **11**, 951-966.

Excoffier, L, Lischer, HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resourses,* **10**, 564-567.

Fernandez-Palacios JM, Nascimento L, Otto R, Delgado JD, Garcia-del-Rey E, Arévalo JR, Wittaker, RJ (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography,* **38**, 226-246.

Fonseca AJAGB (2006) *Study of genetic integrity of* Coturnix coturnix *Linnaeus, 1758 in Portugal and detection of hybridization with an exotic species* Coturnix japonica *Temminck and Schlegel, 1849*. Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity*, **78**, 311-327.

Friesen VL, Congdon BC, Kidd MG, Birt TP (1999) Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. *Molecular Ecology*, **8**, 2147-2149.

GeneCards Online Database[1]. www.genecards.org/cgi-bin/carddisp.pl?gene=GHR. (Accessed: 27th March 2011).

GeneCards Online Database[2]. www.genecards.org/cgi-bin/carddisp.pl?gene=FGB&search=fibrinogen. (Accessed: 27th March 2011).

GeneCards Online Database[3]. www.genecards.org/cgi-bin/carddisp.pl?gene=ACLY&search=acly. (Accessed: 27th March 2011).

Gonzalez J, Castro GD, Garcia-del-Rey E, Berger C, Wink M (2009). Use of mitochondrial and nuclear genes to infer the origin of two endemic pigeons from the Canary Islands. *Journal of Ornithology*, **150**, 357-367.

Grosso AR (2002) *Análise genética do pombo-torcaz* (Columba palumbus): *filogeografia e identificação do pseudogene citocromo B*. Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Oxford University Press, Nucleic Acids Symposium Series No. **41**, 95-98.

Hundsdoerfer AK, Rubinoff D, Attié M, Wink M, Kitching IJ (2009) A revised molecular phylogeny of the globally distributed hawkmoth genus Hyles (Lepidoptera: Sphingidae), based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **52**, 852-865.

Hume R (2002) *Complete Birds of Britain and Europe*. pp 231-234, Dorling Kindersley, London.

Integrated Taxonomic Information System Database. www.itis.gov/index.html. (Accessed: 18th March 2011).

Johnson KP, Clayton DH (2000) Nuclear and Mitochondrial Genes Contain Similar Phylogenetic Signal for Pigeons and Doves (Aves: Columbiformes). *Molecular Phylogenetics and Evolution*, **14**, 141-151.

Johnson KP, Kort S, Dinwoodey K, Mateman AC, Cate CT, Lessells CM, Clayton DH (2001) A molecular phylogeny of the dove genera *Streptopelia* and *Columba*. *The Auk*, **118**, 874-887.

Jourdain E, Gauthier-Clerc M, Bicout DJ, Sabatier P (2007) Bird Migration Routes and Risk for Pathogen Dispersion into Western Mediterranean Wetlands. *Emerging Infectious Diseases*, **13**, 365-372.

Kan XZ, Li XF, Zhang LQ, Chen L, Qian CJ, Zhang XW, Wang L (2010) Characterization of the complete mitochondrial genome of the Rock pigeon, *Columba livia* (Columbiformes: Columbidae). *Genetics and Molecular Research*, **9**, 1234-1249.

Kimball RT, Braun EL, Barker FK, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Han K-L, Harshman J, Heimer-Torres V, Holznagel W, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Reddy S, Sheldon FH, Smith JV, Witt CC, Yuri T (2009) A well-tested set of primers to amplify regions spread across the avian genome. *Molecular Phylogenetics and Evolution*, **50**, 654-660.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Marlier D, Vindevogel H (2006) Viral infections in pigeons. *The Veterinary Journal*, **172**, 40-51.

Martín JL, Cardoso P, Arechavaleta M, Borges PAV, Faria BF, Abreu C, Aguiar AF, Carvalho JA, Costa AC, Cunha RT, Fernandes FM, Gabriel R, Jardim R, Lobo C, Martins AMF, Oliveira P, Rodrigues P, Silva L, Teixeira D, Amorim IR, Homem N, Martins B, Martins M, Mendonça E (2010) Using taxonomically unbiased criteria to prioritize resource allocation for oceanic island species conservation. *Biodiversity and Conservation*, **19**, 1659-1682.

Morin PA, Luikart G, Wayne RK, SNP workshop group (2004) SNPs in ecology, evolution and conservation. *TRENDS in Ecology and Evolution*, **19**, 208-216.

Nichols R (2001) Gene trees and species trees are not the same. *TRENDS in Ecology and Evolution*, **16**, 358-364.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034-1047.

Oliveros CH, Moyle RG (2010) Origin and diversification of Philippine bulbuls. *Molecular Phylogenetics and Evolution*, **54**, 822-832.

Pereira SL, Johnson KP, Clayton DH, Baker AJ (2007) Mitochondrial and Nuclear DNA Sequences Support a Cretaceous Origin of Columbiformes and a Dispersal-Driven Radiation in the Paleogene. *Systematic Biology*, **56**, 656-672.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *TRENDS in Ecology and Evolution*, **24**, 386-393.

Pina-Martins F, Paulo OS (2008) Concatenator: Sequence Data Matrices Handling Made Easy. *Molecular Ecology Resources*, **8**, 1254-1255.

Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**, 603-612.

Prychitko TM, Moore WS (1997) The Utility of DNA Sequences of an Intron from the β-Fibrinogen Gene in Phylogenetic Analysis of Woodpeckers (Aves: Picidae). *Molecular Phylogenetics and Evolution*, **8**, 193-204.

Quintela M, Berlin S, Wang B, Höglund J (2010) Genetic diversity and differentiation among *Lagopus lagopus* populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. *Molecular Ecology*, **19**, 2380-2393.

Ramos-Onsins SE, Rozas J (2002) Statistical Proprieties of New Neutrality Tests Against Population Growth. *Molecular Biology and Evolution*, **19**, 2092-2100.

Rao YS, Wang ZF, Chai XW, Wu GZ, Nie QH, Zhang XQ (2010) Indel segregating within introns in the chicken genome are positively correlated with the recombination rates. *Hereditas*, **147**, 53-57.

Russello MA, Amato G (2004) A molecular phylogeny of *Amazona*: implications for Neotropical parrot biogeography, taxonomy, and conservation. *Molecular Phylogenetics and Evolution*, **30**, 421-437.

Sato A, O'hUigin C, Figueroa F, Grant PR, Grant BR, Tichy H (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proceedings of the National Academy of Sciences*, **96**, 5101-5106.

Schwarz S, Klügel A, Bogaard P, Geldmacher J (2005) Internal structure and evolution of a volcanic rift system in the eastern North Atlantic: the Desertas rift zone, Madeira archipelago. *Journal of volcanology and geothermal research*, **141**, 123-155.

Silva IFR (2007) *Identificação e análise de Polimorfismos Nucleotídicos Simples (SNPs) em* Columba *sp. Linnaeus, 1758: Sua aplicação a uma estratégia de conservação. O pombo como portador de agentes zoonóticos emergentes.* Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*, **136**, 97-107.

Sota T, Vogler AP (2003) Reconstructing species phylogeny of the carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analyses. *Molecular Phylogenetics and Evolution*, **26**, 139-154.

Stephens M, Donnelly P (2003) A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *American Journal of Human Genetics*, **73**, 1162-1169.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876-4882.

UniProt Online Database. www.uniprot.org/uniprot/Q98925. (Accessed: 27th March 2011).

Whittaker RJ, Fernández-Palacios JM (2007) *Island Biogeography – Ecology, evolution and conservation*. Second Edition. Oxford University Press, New York.

Yu L, Zhang Y (2005) Phylogenetic studies of pantherine cats (Felidae) based on multiple genes, with novel application of nuclear β-fibrinogen intron 7 to carnivores. *Molecular Phylogenetics and Evolution*, **35**, 483-495.

Zakharov EV, Lobo NF, Nowak C, Hellmann JJ (2009) Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae). *Heredity*, **102**, 590-599.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563-584.

Zhu L, Zhang Y, Zhang W, Yang S, Chen JQ, Tian D (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics*, **10**, 1-12.

Appendices

Appendix I

Table 7: Samples identification – name, species, origin, sample type and collection date of each sample.

Sample name	Species	Origin	Sample type	Collection date
CtM1	*Columba trocaz*	Madeira	muscle	2005
CtM2	*Columba trocaz*	Madeira	muscle	2005
CtM3	*Columba trocaz*	Madeira	muscle	2005
CtM4	*Columba trocaz*	Madeira	muscle	2005
CtM5	*Columba trocaz*	Madeira	muscle	2005
CtM6	*Columba trocaz*	Madeira	muscle	2010
CtM7	*Columba trocaz*	Madeira	muscle	2010
CtM8	*Columba trocaz*	Madeira	muscle	2010
CtM9	*Columba trocaz*	Madeira	muscle	2010
CtM10	*Columba trocaz*	Madeira	muscle	2010
CtM11	*Columba trocaz*	Madeira	muscle	2010
CtM12	*Columba trocaz*	Madeira	muscle	2010
CtM13	*Columba trocaz*	Madeira	muscle	2010
CtM14	*Columba trocaz*	Madeira	muscle	2010
CtM15	*Columba trocaz*	Madeira	muscle	2010
CpaA20	*Columba palumbus azorica*	Azores - Faial Island	liver	1999
CpaA23	*Columba palumbus azorica*	Azores - Pico Island	liver	1999
CpaA24	*Columba palumbus azorica*	Azores - Pico Island	liver	1999
CpaA25	*Columba palumbus azorica*	Azores - Pico Island	liver	1999
CpaA26	*Columba palumbus azorica*	Azores - Pico Island	liver	1999
CpaA27	*Columba palumbus azorica*	Azores - Pico Island	liver	1999
CpaA29	*Columba palumbus azorica*	Azores - Faial Island	liver	1999
CpaA30	*Columba palumbus azorica*	Azores - Faial Island	liver	1999
CpaA31	*Columba palumbus azorica*	Azores - unkown island	liver	1999
CpaA32	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA33	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA34	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA35	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA36	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA37	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA39	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA40	*Columba palumbus azorica*	Azores - Terceira Island	liver	1999
CpaA41	*Columba palumbus azorica*	Azores - São Miguel Island	liver	1999
CpaA42	*Columba palumbus azorica*	Azores - São Miguel Island	liver	1999
ClaA1	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA2	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA3	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA4	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA5	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA6	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA7	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA8	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA9	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008

ClaA10	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA11	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA12	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA13	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA14	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA15	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA16	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA17	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA18	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA19	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA20	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA21	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA22	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA23	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA24	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA25	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA26	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA27	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA28	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA29	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA30	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA31	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA32	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA33	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA34	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA35	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA36	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA37	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA38	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA39	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA40	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA41	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA42	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA43	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA44	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA45	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA46	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA47	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA48	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA49	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA50	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA51	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA52	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA53	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
ClaA54	*Columba livia atlantis*	Azores - Graciosa Island	feathers	2008
CbC1	*Columba bollii*	Canary Islands	muscle	2005
CbC2	*Columba bollii*	Canary Islands	muscle	2005
CjC	*Columba junoniae*	Canary Islands	muscle	2005
CppFin1	*Columba palumbus palumbus*	Finland	liver	2000
CppFin2	*Columba palumbus palumbus*	Finland	liver	2000
CppFin3	*Columba palumbus palumbus*	Finland	liver	2000

CppFin5	*Columba palumbus palumbus*	Finland	liver	2000
CppFin6	*Columba palumbus palumbus*	Finland	liver	2000
CppFin7	*Columba palumbus palumbus*	Finland	liver	2000
CppFin8	*Columba palumbus palumbus*	Finland	liver	2000
CppFin10	*Columba palumbus palumbus*	Finland	liver	2000
CppFin13	*Columba palumbus palumbus*	Finland	liver	2000
CppFin15	*Columba palumbus palumbus*	Finland	liver	2000
CppFin16	*Columba palumbus palumbus*	Finland	liver	2000
CppFin20	*Columba palumbus palumbus*	Finland	liver	2000
CppFin21	*Columba palumbus palumbus*	Finland	liver	2000
CppFin24	*Columba palumbus palumbus*	Finland	liver	2000
CppFr88	*Columba palumbus palumbus*	France	feathers	2000
CppFr89	*Columba palumbus palumbus*	France	feathers	2000
CppFr96	*Columba palumbus palumbus*	France	feathers	2000
CppFr97	*Columba palumbus palumbus*	France	feathers	2000
CppFr98	*Columba palumbus palumbus*	France	feathers	2000
CppFr99	*Columba palumbus palumbus*	France	feathers	2000
CppFr100	*Columba palumbus palumbus*	France	feathers	2000
CppFr102	*Columba palumbus palumbus*	France	feathers	2000
CppFr107	*Columba palumbus palumbus*	France	feathers	2000
CppFr109	*Columba palumbus palumbus*	France	feathers	2000
CppFr110	*Columba palumbus palumbus*	France	feathers	2000
CppFr113	*Columba palumbus palumbus*	France	feathers	2000
CppFr114	*Columba palumbus palumbus*	France	feathers	2000
CppG6	*Columba palumbus palumbus*	Germany	liver	2000
CppG7	*Columba palumbus palumbus*	Germany	liver	2000
CppG9	*Columba palumbus palumbus*	Germany	liver	2000
CppH1	*Columba palumbus palumbus*	Hungary	liver	2000
CppH2	*Columba palumbus palumbus*	Hungary	liver	2000
CppH3	*Columba palumbus palumbus*	Hungary	liver	2000
CppH4	*Columba palumbus palumbus*	Hungary	liver	2000
CppH5	*Columba palumbus palumbus*	Hungary	liver	2000
CppH6	*Columba palumbus palumbus*	Hungary	liver	2000
CppH7	*Columba palumbus palumbus*	Hungary	liver	2000
CppH10	*Columba palumbus palumbus*	Hungary	liver	2000
CppH11	*Columba palumbus palumbus*	Hungary	liver	2000
CppH13	*Columba palumbus palumbus*	Hungary	liver	2000
CppH15	*Columba palumbus palumbus*	Hungary	liver	2000
CppH23	*Columba palumbus palumbus*	Hungary	liver	2000
CppH30	*Columba palumbus palumbus*	Hungary	liver	2000
CppLa1	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa2	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa3	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa4	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa5	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa6	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa7	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa8	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa9	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa10	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa11	*Columba palumbus palumbus*	Latvia	liver	2000

CppLa12	*Columba palumbus palumbus*	Latvia	liver	2000
CppLa13	*Columba palumbus palumbus*	Latvia	liver	2000
CppLit1	*Columba palumbus palumbus*	Lithuania	liver	2000
CppLit2	*Columba palumbus palumbus*	Lithuania	liver	2000
CppLit3	*Columba palumbus palumbus*	Lithuania	liver	2000
CppPol1	*Columba palumbus palumbus*	Poland	liver	2000
CppPol2	*Columba palumbus palumbus*	Poland	liver	2000
CppPol5	*Columba palumbus palumbus*	Poland	liver	2000
CppPt1	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt2	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt3	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt4	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt5	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt6	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt7	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt8	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt9	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt10	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt11	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt12	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt13	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt14	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt15	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt16	*Columba palumbus palumbus*	Portugal	liver	2000
CppPt17	*Columba palumbus palumbus*	Portugal	liver	2000
CppSp58	*Columba palumbus palumbus*	Spain	feathers	2000
CppSp80	*Columba palumbus palumbus*	Spain	feathers	2000
CppSp85	*Columba palumbus palumbus*	Spain	feathers	2000
CppSw1	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw2	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw3	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw4	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw5	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw6	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw7	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw8	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw10	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw14	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw16	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw17	*Columba palumbus palumbus*	Sweden	liver	2000
CppSw18	*Columba palumbus palumbus*	Sweden	liver	2000

Appendix II

Table 8: Amplified samples for each fragment in the present work (dark grey) or previously (light grey) in Silva (2007).

	Gene					
Sample	**GHR**	**RP40**	**TROP**	**β-FIB**	**TGF-β2**	**IRF2**
CtM1	▒	▒	▒		■	■
CtM2	▒	▒	▒	▒	■	■
CtM3	▒	▒	▒		■	■
CtM4	■	■	■	■	■	■
CtM5	▒	■	■	■	■	■
CtM6	■	■	■	■	■	
CtM7	■	■	■	■	■	
CtM8	■	■	■	■	■	■
CtM9	■	■		■	■	
CtM10	■	■	■	■	■	
CtM11	■	■		■	■	
CtM12	■	■	■	■	■	
CtM13	■	■	■	■	■	
CtM14	■	■	■	■	■	■
CtM15	■	■	■	■	■	■
CpaA20	■	■	■	■	■	■
CpaA23	▒	▒	▒			
CpaA24	▒	■	■	■	■	
CpaA25	▒	■	■	■	■	
CpaA26	▒	■	▒	▒	■	
CpaA27	▒	▒	▒		■	■
CpaA29	▒	■		■	■	■
CpaA30	■	■			■	
CpaA31	▒	■	▒	■	■	■
CpaA32	■	■	■	■	■	■
CpaA33	■	■			■	
CpaA34	■	■			■	
CpaA35	■	■			■	■
CpaA36	■	■	■	■	■	■
CpaA37	▒	■			■	■
CpaA39	■	■		■	■	■
CpaA40	■	■		▒	■	■
CpaA41	■	■				■
CpaA42	▒	■		■	■	■
CbC1	▒		▒	▒		
CbC2	▒	▒	▒			

CjC	░	░	░	░		
CppFin1	░	░				
CppFin2	░					
CppFin3	░	░	░			
CppFin5	░	░	░			
CppFin6			░			
CppFin7	░	█				
CppFin8	░					
CppFin10	█	█	█	█	█	█
CppFin13	█	█				
CppFin15	░					
CppFin16	█	█				
CppFin20	░					
CppFin21	░					
CppFin24	░					
CppFr88	█	█				
CppFr89	░					
CppFr96	░	█				
CppFr97	█	█		█	█	
CppFr98	░	█				
CppFr99	█	█				
CppFr100	░					
CppFr102	░					
CppFr107	░					
CppFr109	░					
CppFr110	░					
CppFr113	░					
CppFr114	░	█				
CppG6	█	█				
CppG7		█				
CppG9	█	█	█	█	█	
CppH1	░	░	░	░		
CppH2	░	░	░	░		
CppH3	█	█			█	
CppH4	█	█				
CppH5	█	█			█	
CppH6	░	░	░	░		
CppH7	░					
CppH10	░			░		
CppH11	░					
CppH13	░					
CppH15	░					
CppH23	░					
CppH30	░					
CppLa1	░					
CppLa2		█				

CppLa3	░					
CppLa4	░					
CppLa5	■	■				
CppLa6	░					
CppLa7	░					
CppLa8	░					
CppLa9	░					
CppLa10	░					
CppLa11	░					
CppLa12	■	■			■	
CppLa13	░					
CppLit1	■	■				
CppLit2	■	■				
CppLit3	■				■	
CppPol1	■	■				
CppPol2	■					
CppPol5	■	■	■	■	■	
CppPt1	░					
CppPt2	░	░	░	░		
CppPt3	░	░	░	░		
CppPt4	■	░	░	░		
CppPt5	░					
CppPt6	░	■				
CppPt7	■	■				
CppPt8	░	■				
CppPt9	░					
CppPt10	■	■				
CppPt11	■	■	■	■	■	
CppPt12	■	■		■	■	■
CppPt13	■	■				
CppPt14	■	■				
CppPt15	■	■				
CppPt16	■	■				
CppPt17	░	■			■	■
CppSp58	■	■	■	■	■	■
CppSp80	■	■				
CppSp85	■	■				
CppSw1	■	■				
CppSw2	░					
CppSw3	░					
CppSw4	■	■				
CppSw5	░					
CppSw6	░					
CppSw7	░					
CppSw8	■	■	■	■	■	

CppSw10	░					
CppSw14	░					
CppSw16	░					
CppSw17	░					
CppSw18	░	▓				

Appendix III

Phenol / Chloroform Extraction Protocol

Day 1

1) Prepare a 2 ml tube with 500 µl of extraction buffer (TrisHCl, EDTA, NaCl, SDS and water) + 10 µl de proteinase K (10 mg/ml);
2) Put the tissue sample (muscle, liver or feather) inside the tube;
3) Incubate the tubes overnight at 50ºC

Day 2

4) Add 150 µl of saturated NaCl;
5) Mix rigorously for about 15 seconds;
6) Add 1 volume of phenol / chloroform 25:24 (650 µl for each 2 ml tube);
7) Mix gently by inversion for 5 minutes;
8) Centrifuge 5 minutes (9000 rpm);
9) Remove the aqueous phase;
10) Repeat the steps 6, 7, 8 and 9;
11) Add 1 volume (650 µl) of chloroform / isoamilic acid (24:1);
12) Mix gently by inversion for 5 minutes;
13) Centrifuge 5 minutes (9000 rpm);
14) Remove the aqueous phase;
15) Add 2.5 volumes (about 1500 µl) of absolute ethanol;
16) Mix gently by inversion
17) Put the tubes at -80ºC for 18 minutes or at -20ºC for 1 hour to precipitate DNA;
18) Centrifuge 15 min at 4ºC (13000 rpm)
19) Discard the supernatant;
20) Add 500 µl of 70% ethanol;
21) Centrifuge 10 min at 4ºC (13000 rpm)
22) Discard the supernatant;
23) Dry the pellet (overnight at room temperature; 15-30 minutes at 37ºC; 3-5 minutes at 55ºC);
24) Resuspend the DNA in 100 µl in water for one day.

Appendix IV

Table 9: Used samples in the concatenated dataset. White spaces mean that those samples were not amplified for that fragment.

	Gene				
Sample	**GHR**	**RP40**	**TROP**	**β-FIB**	**TGB-β2**
CtM1	■	■	■		■
CtM2	■	■	■	■	■
CtM3	■	■	■		■
CtM4-1	■	■	■	■	■
CtM4-2	■	■	■	■	■
CtM5-1	■	■	■	■	■
CtM5-2	■	■	■	■	■
CtM6	■	■	■	■	■
CtM7-1	■	■	■	■	■
CtM7-2	■	■	■	■	■
CtM8	■	■	■	■	■
CtM9	■	■		■	■
CtM10	■	■	■	■	■
CtM11	■	■		■	■
CtM12	■	■	■	■	■
CtM13-1	■	■	■	■	■
CtM13-2	■	■	■	■	■
CtM14	■	■	■	■	■
CtM15	■	■	■	■	■
CpaA20-1	■	■	■	■	■
CpaA20-2	■	■	■	■	■
CpaA24	■	■	■	■	■
CpaA25-1	■	■	■	■	■
CpaA25-2	■	■	■	■	■
CpaA26	■	■	■	■	■
CpaA27	■	■	■		■
CpaA29-1	■	■		■	■
CpaA29-2	■	■		■	■
CpaA31	■	■	■	■	■
CpaA32	■	■	■	■	■
CpaA36	■	■	■	■	■
CpaA39-1	■	■		■	■
CpaA39-2	■	■		■	■
CpaA40	■	■		■	■
CpaA42	■	■		■	■
CjC	■	■	■	■	
CppFin10-1	■	■	■	■	■
CppFin10-2	■	■	■	■	■
CppFr97	■	■		■	■

CppG9-1	■	■	■	■	■
CppG9-2	■	■	■	■	■
CppH1	■	■	■	■	
CppH2	■	■	■	■	
CppH6	■	■	■		■
CppPol5-1	■	■	■	■	■
CppPol5-2	■	■	■	■	■
CppPt2	■	■	■	■	
CppPt3	■	■	■	■	
CppPt4	■	■	■	■	
CppPt11-1	■	■	■	■	■
CppPt11-2	■	■	■	■	■
CppPt12	■	■		■	■
CppSp58	■	■	■	■	■
CppSw8	■	■	■	■	■

Appendix V

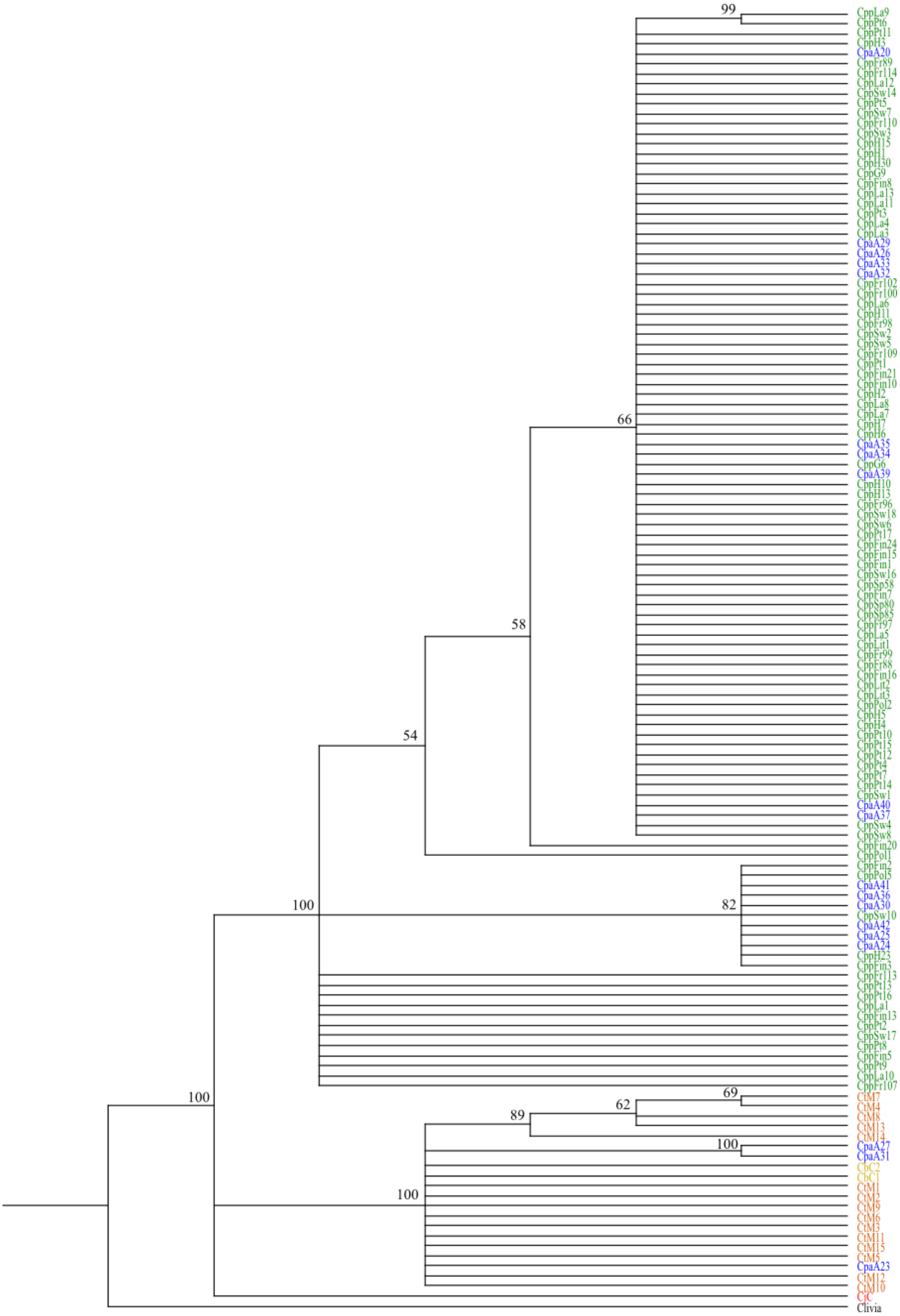

Figure 28: Consensus tree obtained from the 1000 trees on MP analysis of GHR intron 9 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

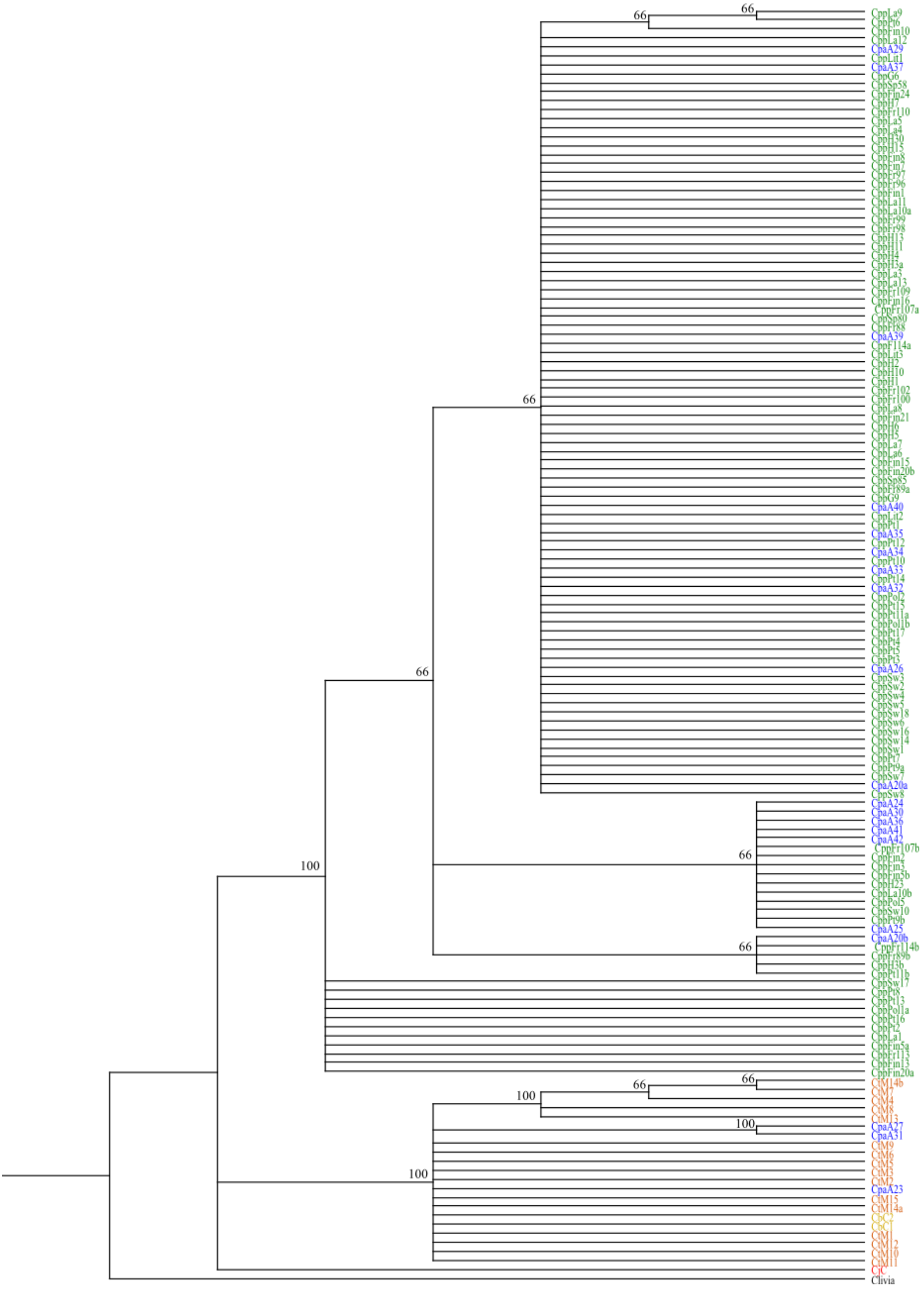

Figure 29: Consensus tree obtained from the 54 trees on MP analysis of GHR intron 9 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

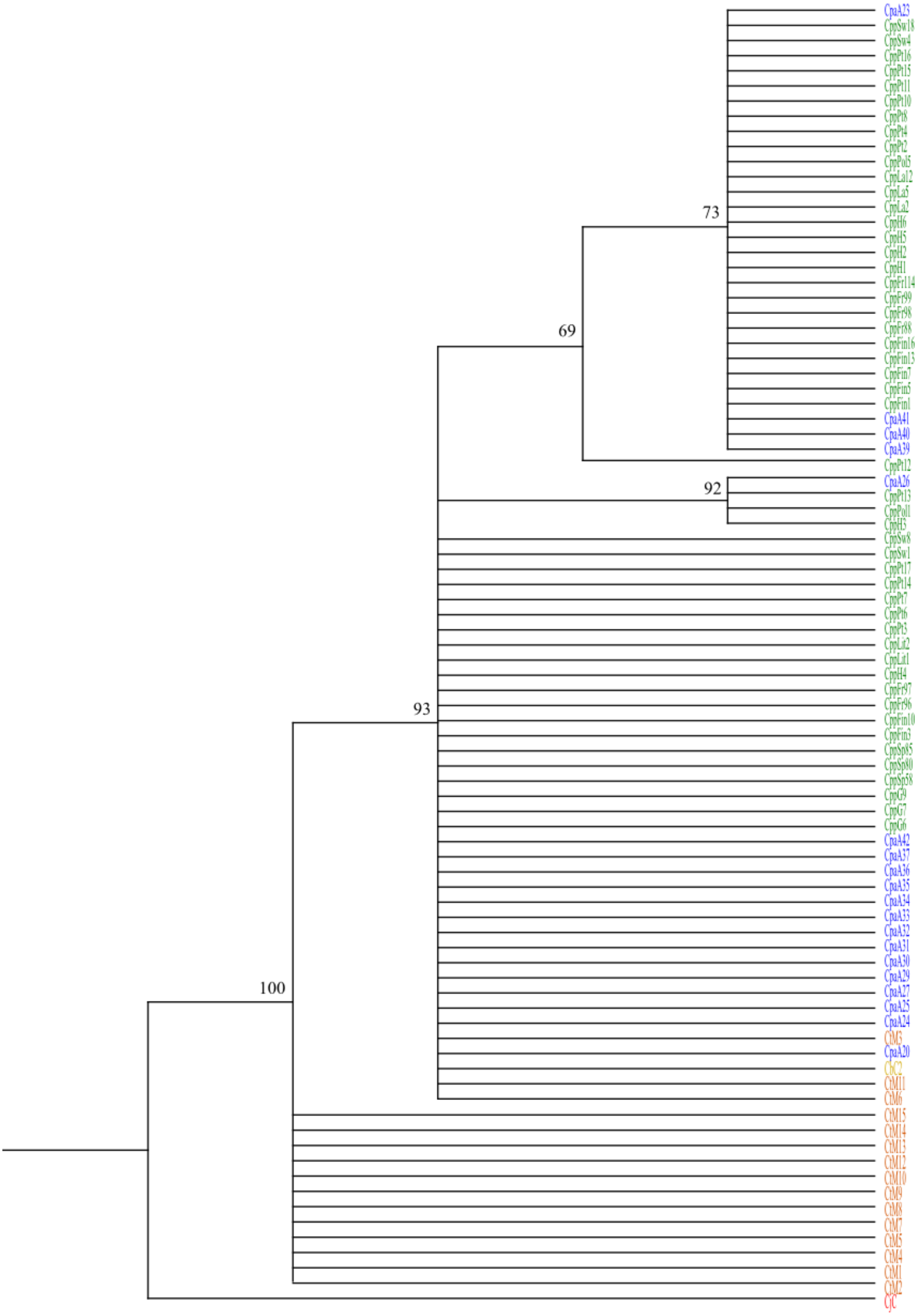

Figure 30: Consensus tree obtained from the 1000 trees on MP analysis of RP40 intron 5 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

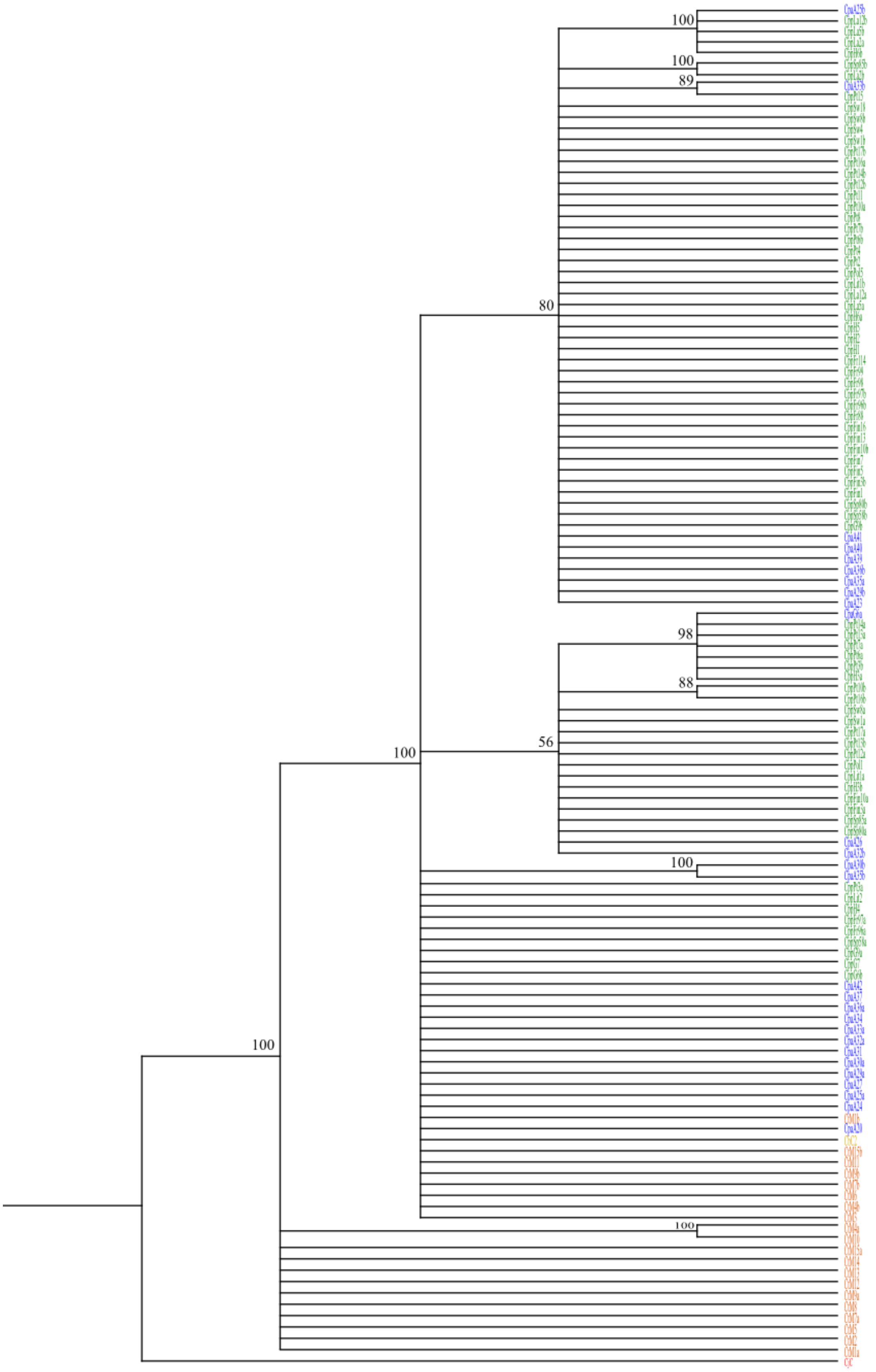

Figure 31: Consensus tree obtained from the 1000 trees on MP analysis of RP40 intron 5 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

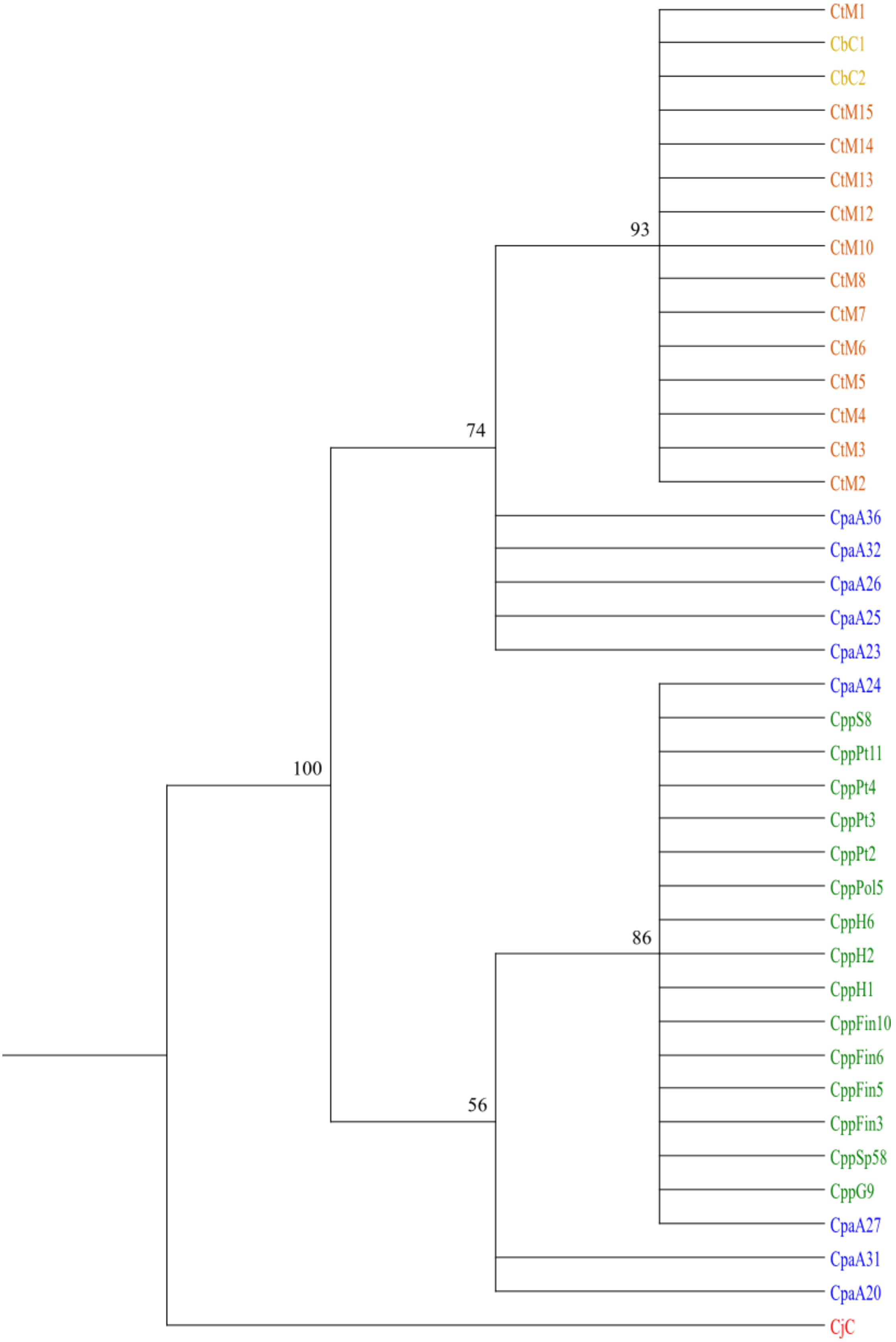

Figure 32: Consensus tree obtained from the 1000 trees on MP analysis of TROP intron 5 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*.

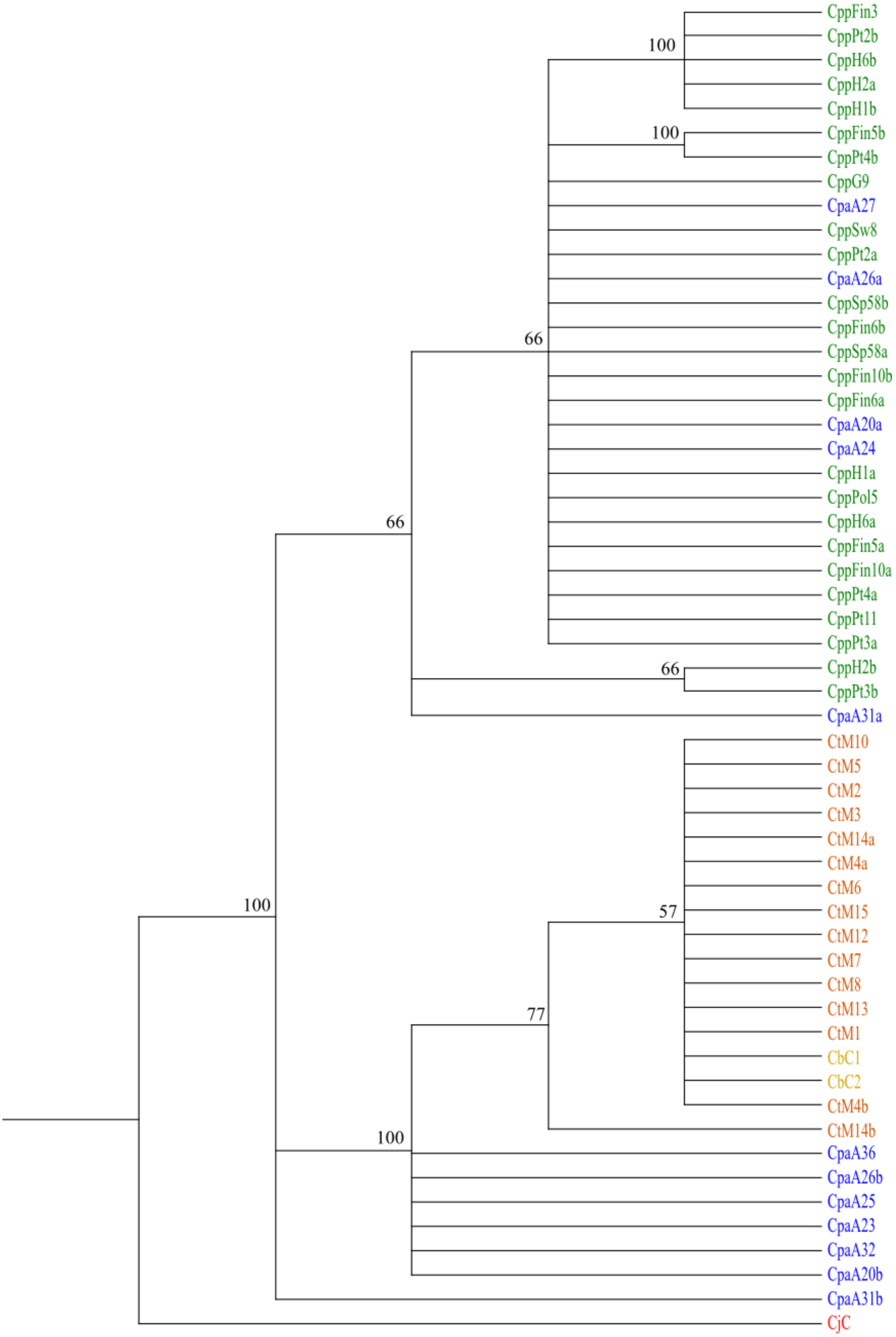

Figure 33: Consensus tree obtained from the 270 trees on MP analysis of TROP intron 5 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

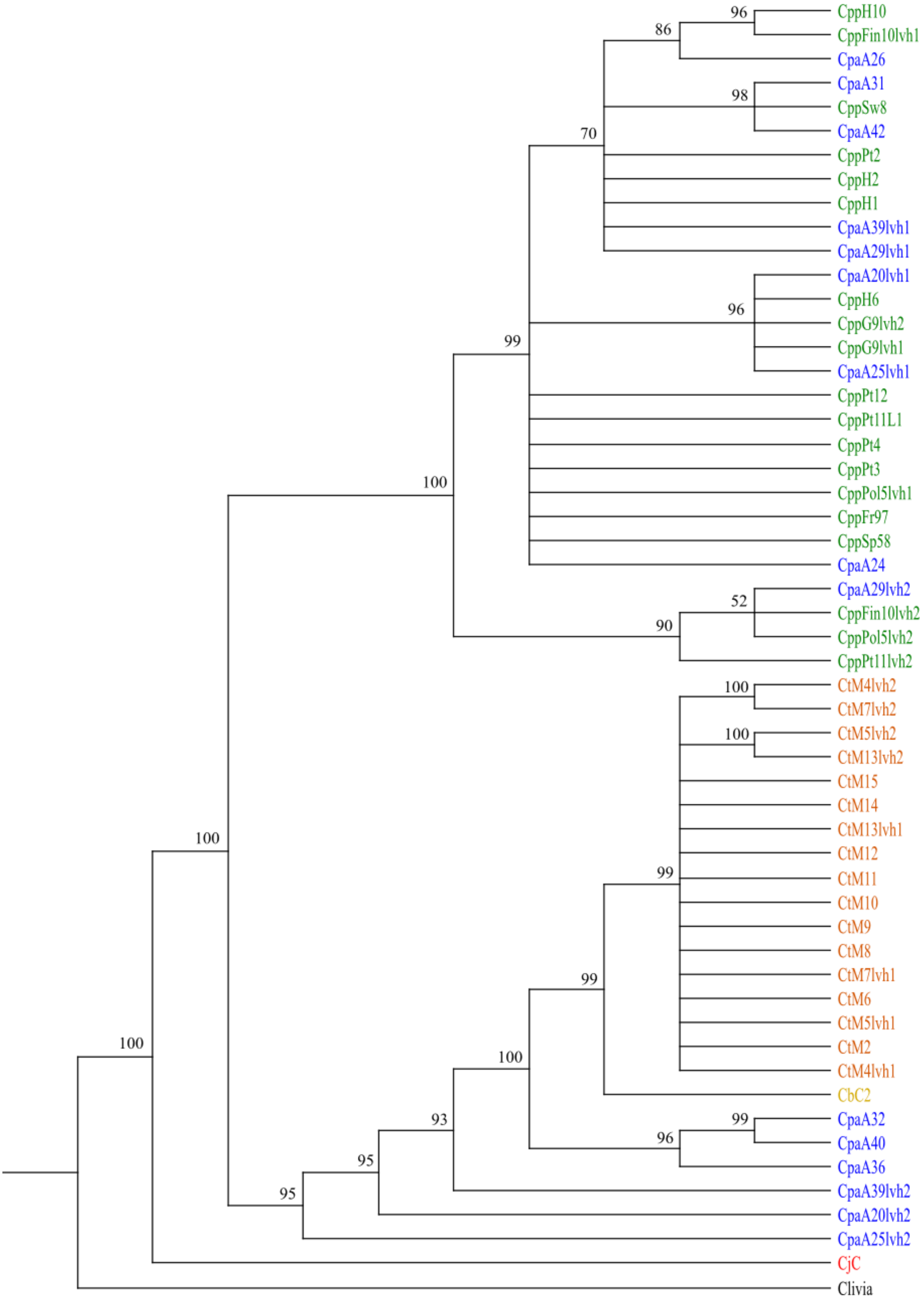

Figure 34: Consensus tree obtained from the 1000 trees on MP analysis of β-FIB intron 7 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. *lvh1* and *lvh2* refer to the both possible haplotypes when the sample has a length variation in the fragment.

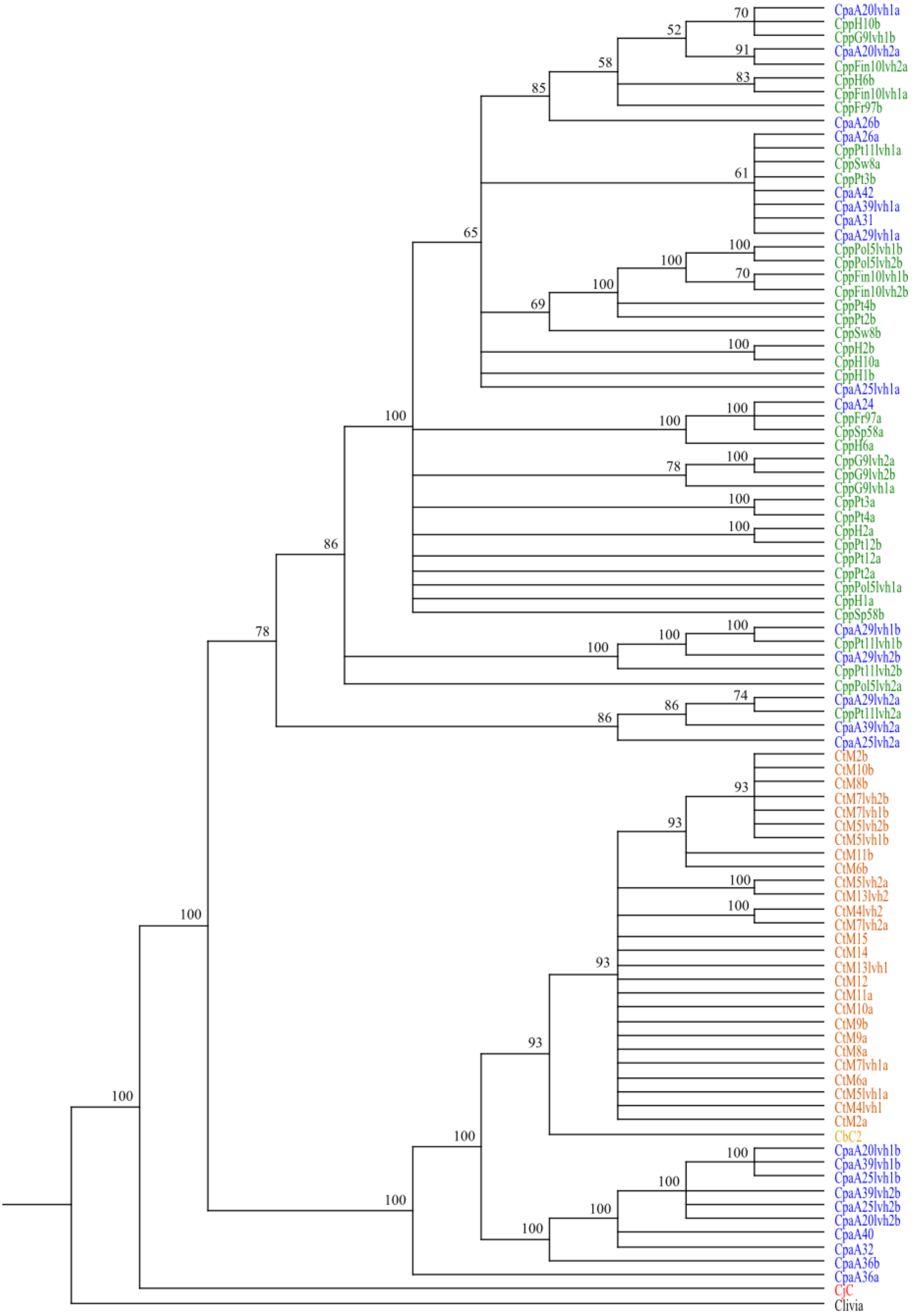

Figure 35: Consensus tree obtained from the 1000 trees on MP analysis of β-FIB intron 7 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. The samples with both heterozygous sites and length variation, four possible options are shown - *lvh1a, lvh1b, lvh2a and lvh2b*.

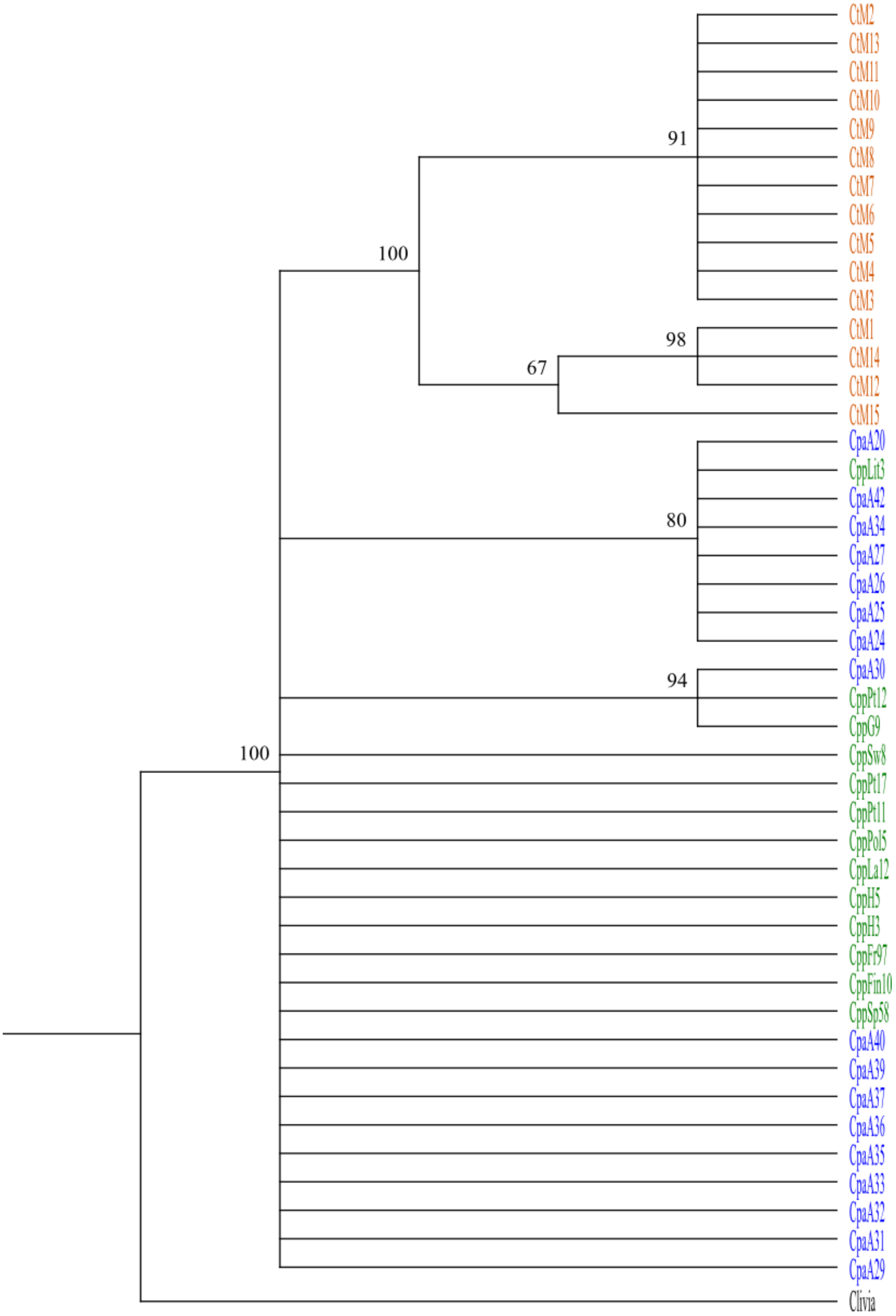

Figure 36: Consensus tree obtained from the 1000 trees on MP analysis of TGF-β2 intron 5 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*.

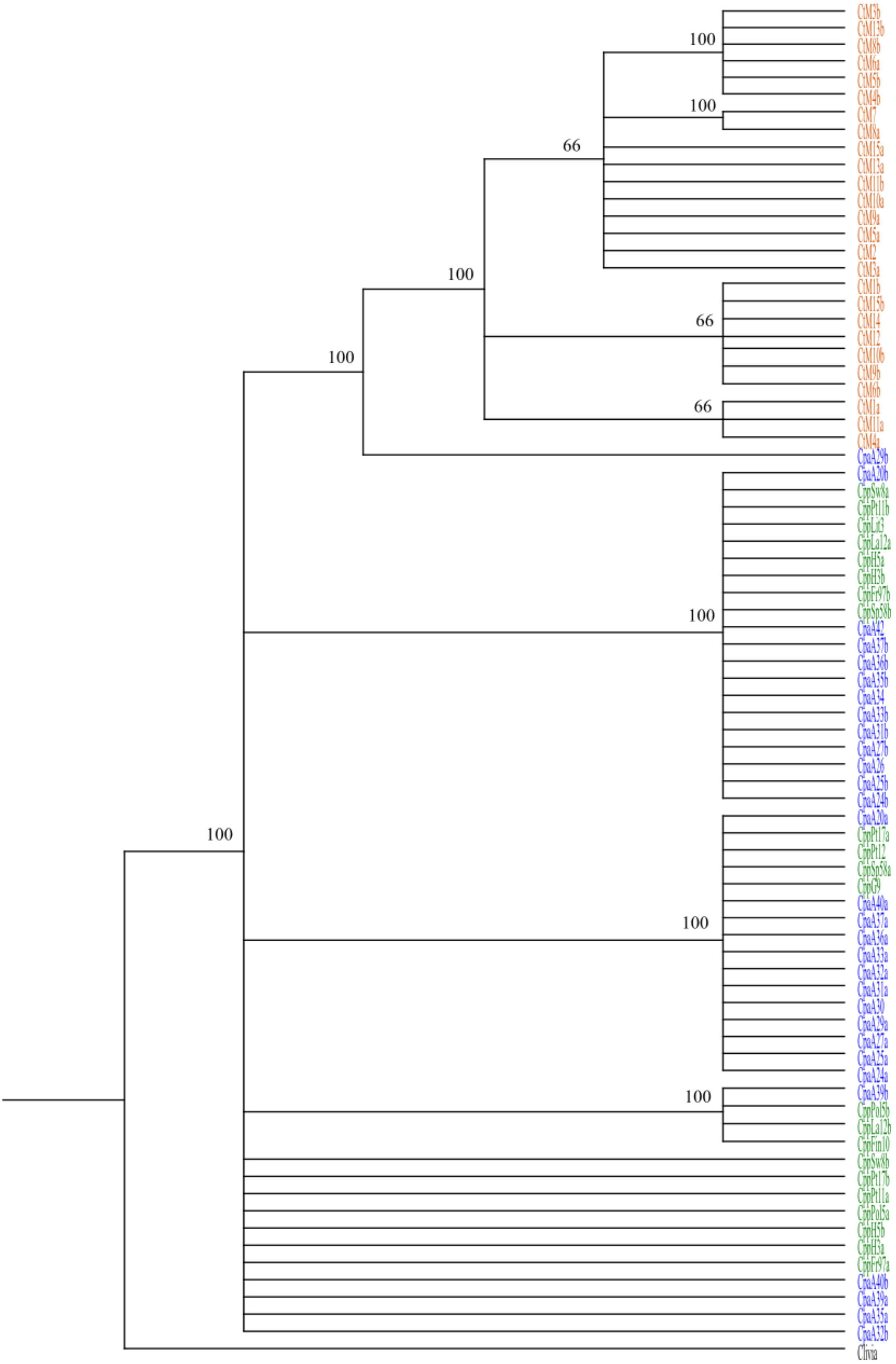

Figure 37: Consensus tree obtained from the 30 trees on MP analysis of TGF-β2 intron 5 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

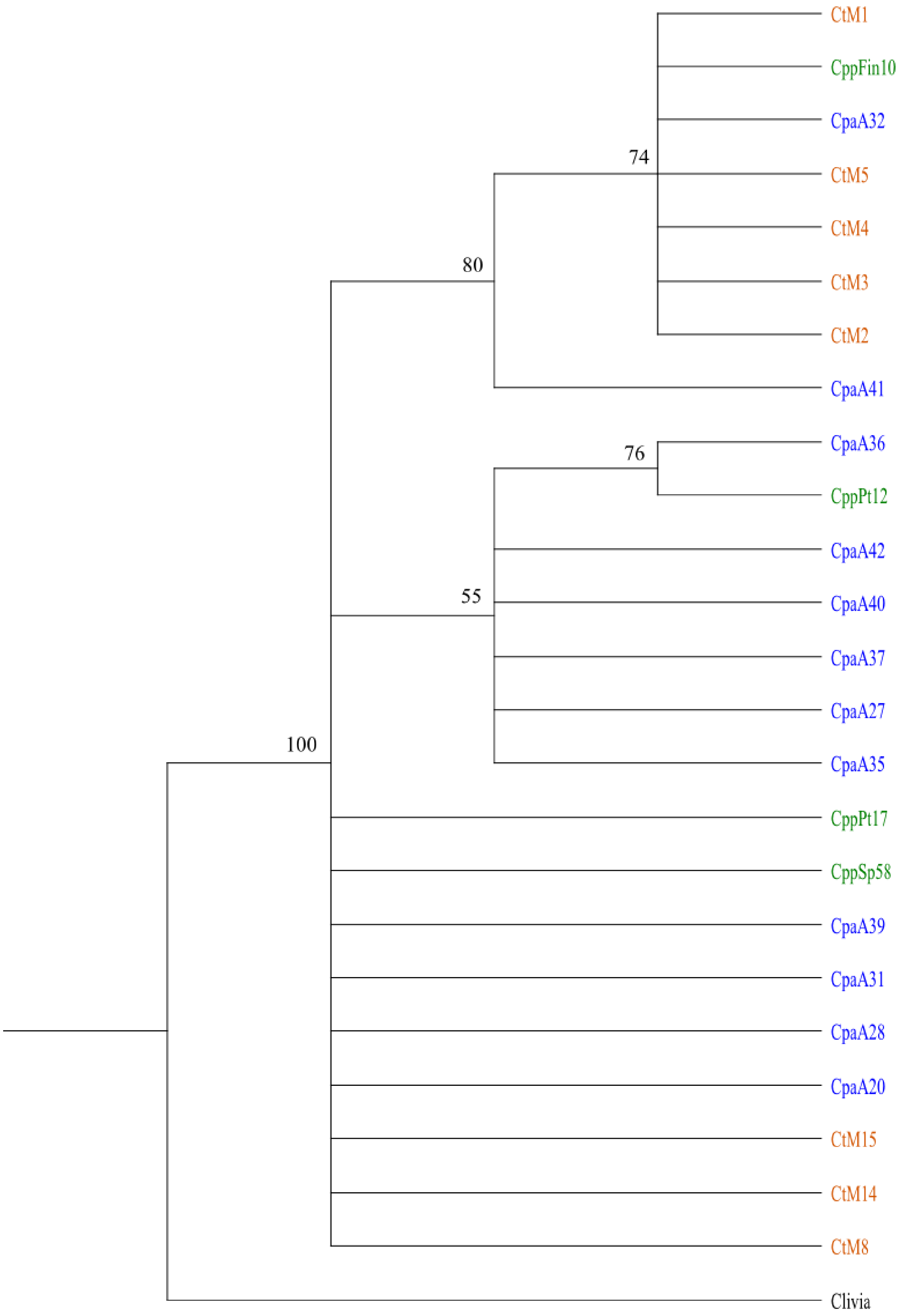

Figure 38: Consensus tree obtained from the 1000 trees on MP analysis of IRF2 intron 2 sequence data with heterozygous bases dealt as ambiguities. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*.

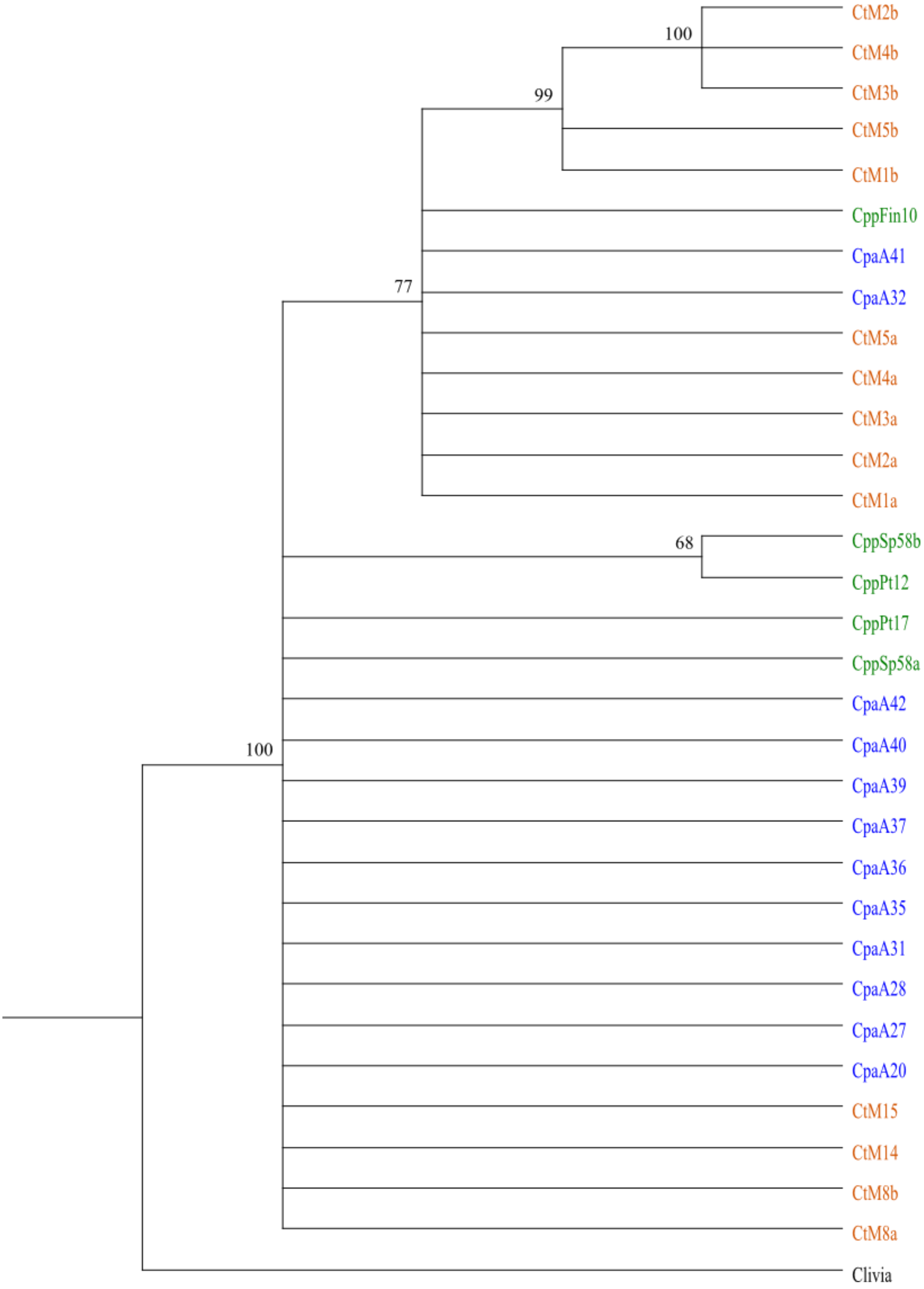

Figure 39: Consensus tree obtained from the 1000 trees on MP analysis of IRF2 intron 2 haplotypic data. Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. *a* and *b* refer to both haplotypes of a sample with ambiguous bases in this fragment.

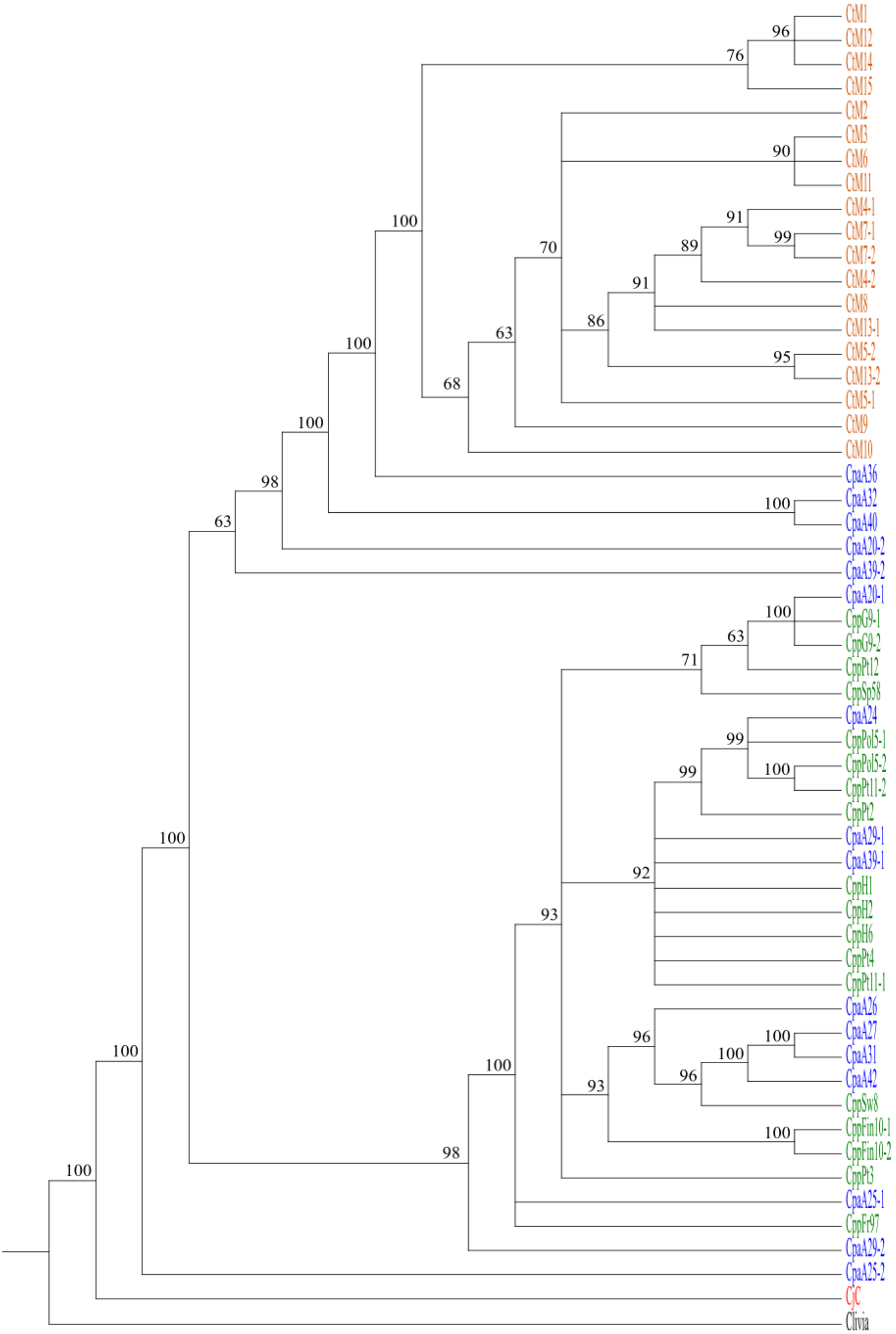

Figure 40: Consensus tree obtained from the 1000 trees on MP analysis of 5 concatenated fragments (GHR, RP40, TROP, β-FIB and TGF-β2). Numbers indicate consensus values above 50%. Colors indicate the species: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz* and red – *C. junoniae*. . *1* and *2* refer to both concatenation options due to the presence of a length variation in β-FIB intron 7.

Appendix VI

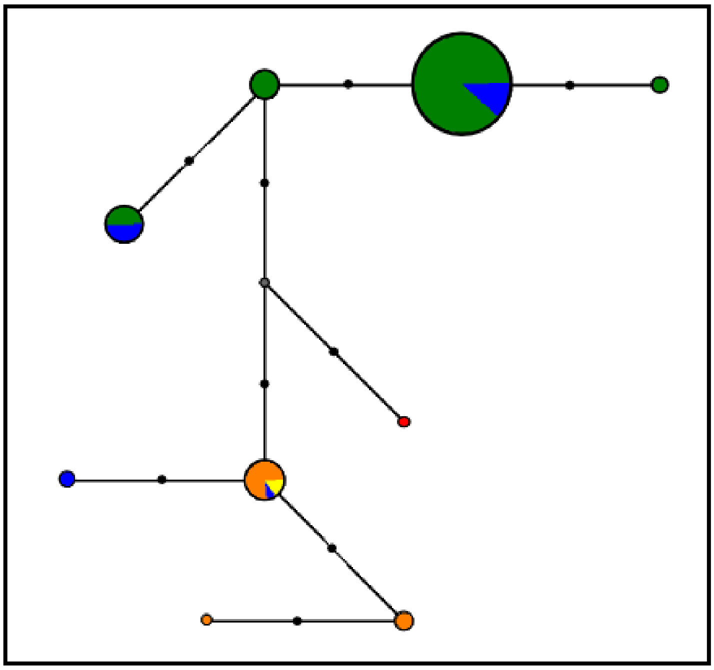

Figure 41: Median-joining network from GHR intron 9 data with heterozygous bases dealt as ambiguities. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

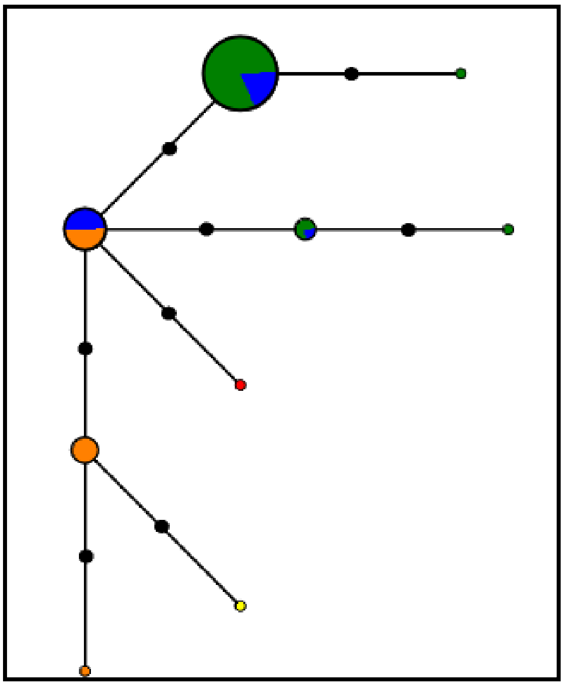

Figure 42: Median-joining network from RP40 intron 5 data with heterozygous bases dealt as ambiguities. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. Black circles correspond to mutation events.

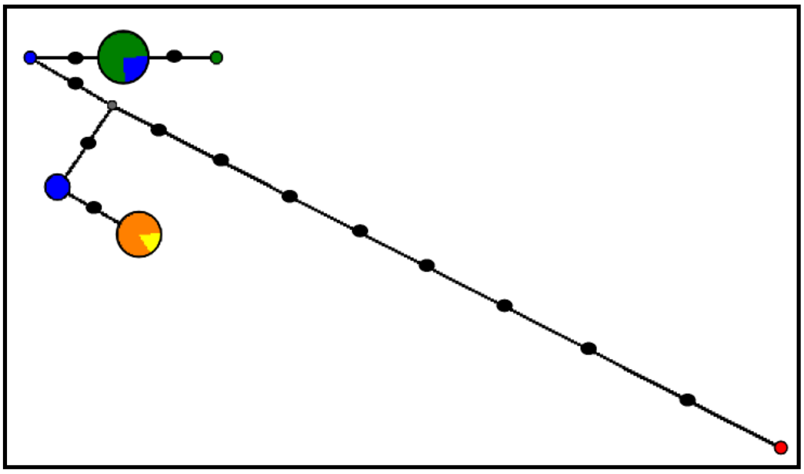

Figure 43: Median-joining network from TROP intron 5 data with heterozygous bases dealt as ambiguities. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

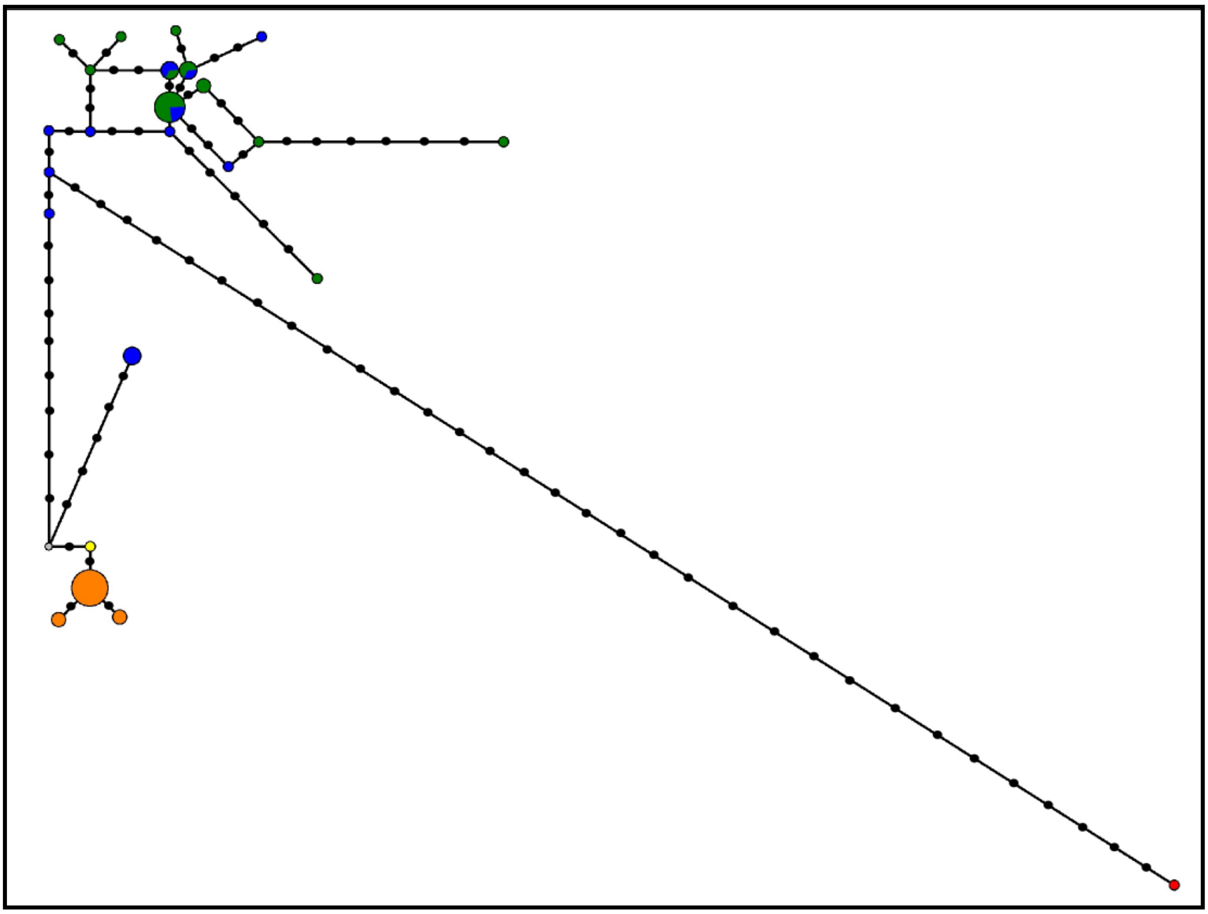

Figure 44: Median-joining network from β-FIB intron 7 data with heterozygous bases dealt as ambiguities. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica*; orange – *C. trocaz*; yellow – *C. bollii* and red – *C. junoniae*. A median vector is represented by a grey circle. Black circles correspond to mutation events.

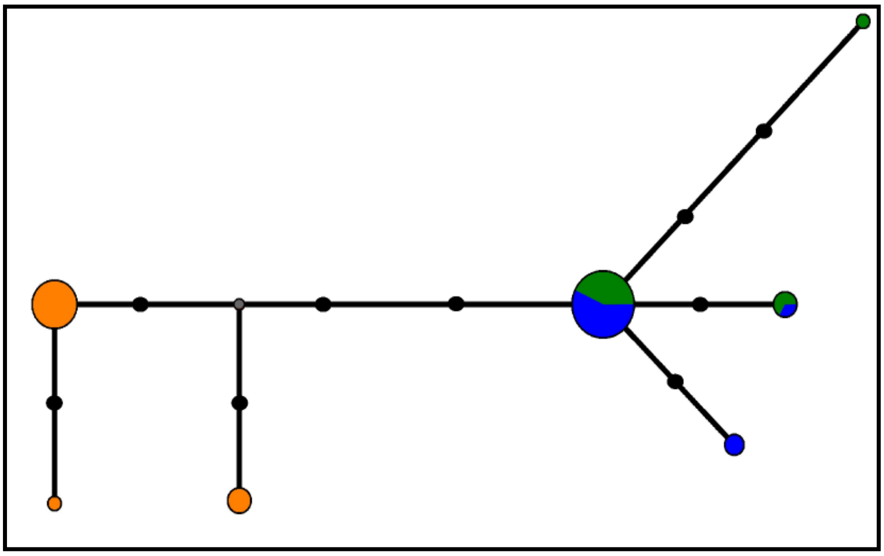

Figure 45: Median-joining network from TGF-β2 intron 5 data with heterozygous bases dealt as ambiguities. Circle size is proportional to the frequency of each haplotype. Each species is represented by a different color: green – *C. p. palumbus*; blue – *C. p. azorica* and orange – *C. trocaz*. A median vector is represented by a grey circle. Black circles correspond to mutation events.