



DR. NICOLAS DUSSEX (Orcid ID : 0000-0002-9179-8593)

DR. VERENA ESTHER KUTSCHERA (Orcid ID : 0000-0002-8930-534X)

PROF. NEIL JOHN GEMMELL (Orcid ID : 0000-0003-0671-3637)

Article type : Original Article

## **A genome-wide investigation of adaptive signatures in protein-coding genes related to tool behaviour in New Caledonian and Hawaiian crows**

Nicolas Dussex<sup>1,2\*</sup>, Verena E. Kutschera<sup>3,4\*</sup>, R. Axel W. Wiberg<sup>5,6\*</sup>, Darren J. Parker<sup>5,7</sup>, Gavin R. Hunt<sup>8†</sup>, Russell D. Gray<sup>8,9</sup>, Kim Rutherford<sup>1</sup>, Hideaki Abe<sup>1,10</sup>, Robert C. Fleischer<sup>11</sup>, Michael G. Ritchie<sup>5§</sup>, Christian Rutz<sup>5§†</sup>, Jochen B. Wolf<sup>3,12§</sup>, Neil J. Gemmell<sup>1§</sup>

<sup>1</sup>*Department of Anatomy, University of Otago, PO Box 913, Dunedin 9016, New Zealand*

<sup>2</sup>*Centre for Palaeogenetics, Svante Arrhenius väg 20C, Stockholm SE-10691, Sweden*

<sup>3</sup>*Department of Evolutionary Biology, Norbyvägen 18D, Uppsala University, Uppsala 75236, Sweden*

<sup>4</sup>*Dept of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Box 1031, SE-17121 Solna, Sweden*

<sup>5</sup>*Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews KY16 9TH, UK*

<sup>6</sup>*Department of Environmental Sciences, Evolutionary Biology, University of Basel, Basel, Switzerland*

<sup>7</sup>*Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne 1015, Switzerland.*

<sup>8</sup>*University of Auckland, Science Centre 302, 23 Symonds st, Auckland 1010, New Zealand*

<sup>9</sup>*Max Planck Institute for the Science of Human History, Kahlaische Strasse 10, Jena D-07745, Germany*

<sup>10</sup>*Wildlife Research Center, Kyoto University, Kyoto, Japan*

<sup>11</sup>*Center for Conservation Genomics, Smithsonian Conservation Biology Institute, Washington, DC 20008, USA*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/MEC.15775](https://doi.org/10.1111/MEC.15775)

This article is protected by copyright. All rights reserved

<sup>12</sup>Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

\*These authors contributed equally

§These authors' laboratories contributed equally

GRH: Unaffiliated.

Correspondence: nicolas.dusseix@gmail.com; Centre for Palaeogenetics, Svante Arrhenius väg 20C, Stockholm SE-10691, Sweden; phone: 0046(0)700317026

*Keywords: corvids, tool use, selection, genomics*

## **Abstract**

Very few animals habitually manufacture and use tools. It has been suggested that advanced tool behaviour co-evolves with a suite of behavioural, morphological and life-history traits. In fact, there are indications for such an adaptive complex in tool-using crows (genus *Corvus* species). Here, we sequenced the genomes of two habitually tool-using and ten non-tool-using crow species to search for genomic signatures associated with a tool-using lifestyle. Using comparative genomic and population genetic approaches, we screened for signals of selection in protein-coding genes in the tool-using New Caledonian and Hawaiian crows. While we detected signals of recent selection in New Caledonian crows near genes associated with bill morphology, our data indicate that genetic changes in these two lineages are surprisingly subtle, with little evidence at present for convergence. We explore the biological explanations for these findings, such as the relative roles of gene regulation and protein-coding changes, as well as the possibility that statistical power to detect selection in recently diverged lineages may have been insufficient. Our study contributes to a growing body of literature aiming to decipher the genetic basis of recently evolved complex behaviour.

## **Introduction**

Humans' extraordinary ability to make and use tools is believed to have co-evolved with a suite of unusual traits, including: a large brain, advanced cognitive capacities, language, bipedalism, opposable thumbs, an extended developmental period, extreme sociality, and an ability to culturally transmit and accumulate knowledge (Henrich, 2015; Kaplan, Gurven, & Winking, 2009; Sprevak, 2014; Sterelny, 2012). Inferring causal inter-relationships between these traits is extremely challenging, not least because some of them have reached degrees of sophistication or specialisation not found in any other species (Gibson, Gibson, & Ingold, 1994). But perhaps other tool users have evolved comparable tool-use-related traits? If so, what might their underlying genetic architectures be and can these be identified using genomic approaches? Addressing this challenging topic requires comparisons between multiple tool-using and non-tool-using species, ideally from a non-primate system (Rutz et al., 2016). Given the rarity of animal tool behaviour (Biro, Haslam, & Rutz, 2013; Hunt, Gray, & Taylor, 2010; Shumaker, Walkup, & Beck, 2011), the genus *Corvus*, which comprises over 40 species of crows and ravens (Haring, Däubel, Pinsker, Kryukov, & Gamauf, 2012), including at least two habitual tool users (Rutz et al., 2016), offers an exceptional opportunity for such replicated analyses.

The New Caledonian crow (*Corvus moneduloides*), from the South Pacific archipelago of New Caledonia, is one of the most accomplished non-human tool users (Hunt, 1996; Hunt & Gray, 2002, 2003). The species is able to forage with, as well as manufacture, a diverse repertoire of tools (Hunt, 1996; Hunt & Gray, 2002, 2003; Rutz & St Clair, 2012; St Clair et al., 2018) and exhibits a range of traits that appear to be related to its tool-using lifestyle (Hunt & Uomini 2016). These include: (i) a genus-atypical bill morphology that aids dexterous tool handling (Matsui et al., 2016; Troscianko, von Bayern, Chappell, Rutz, & Martin, 2012); (ii) large, mobile eyes and an exceptionally wide field of binocular vision facilitating visual feedback and motor coordination during tool manipulation (Matsui et al., 2016; Troscianko et al., 2012); (iii) genetic predispositions that enable the ontogenetic development of basic tool skills (Kenward, Rutz, Weir, & Kacelnik, 2006; Kenward, Weir, Rutz, & Kacelnik, 2005); (iv) a prolonged juvenile dependency period providing opportunities for juveniles to hone their technical skills and learn from proficient adults (Holzhaider et al. 2011; Hunt et al. 2012; Bluff et al. 2010); and (v) a highly dynamic social system that may aid the social transmission of tool-related skills within and between generations (Holzhaider et al., 2011; St

Clair et al., 2015).

It has also been reported that New Caledonian crows have unusually large brains (Cnotka, Güntürkün, Rehkämper, Gray, & Hunt, 2008) with relatively enlarged associative forebrain regions (Mehlhorn, Hunt, Gray, Rehkämper, & Güntürkün, 2010), but these results could not be replicated with a larger sample of birds (Medina, 2013). While a number of studies have argued that the species possesses impressive problem-solving abilities (Gruber et al., 2019; Taylor, Elliffe, Hunt, & Gray, 2010), some claims have been contested (for a recent review, see Rutz & Hunt 2020) and evidence that New Caledonian crows' cognitive abilities exceed those of other corvids remains equivocal (see Taylor and Gray 2014). Indeed, very few studies have been carried out to date that compare the cognitive performance of New Caledonian crows to that of non-tool-using congeners, using tasks that do not require object manipulation (but see Teschke et al. (2013)).

While some non-tool-using crow species can learn to use tools in captivity, usually after some training or facilitation (Bird & Emery, 2009; Gallot & Gruber, 2019; Kanai, Matsui, Watanabe, & Izawa, 2014; Powell & Kelly, 1977), it was long thought that the New Caledonian crow is the only natural tool user within the genus. Recent work revealed, however, that the critically-endangered Hawaiian crow (*C. hawaiiensis*), from Hawai'i Island in the North Pacific, is also highly skilled at making and using tools (Klump, Masuda, St Clair, & Rutz, 2018; Rutz et al., 2016). Two lines of evidence suggest that tool use is part of the Hawaiian crow's natural behavioural repertoire (Rutz et al., 2016): naïve juveniles spontaneously develop functional tool use (i.e., without demonstrations or training), and tool use is a species-wide capacity (adults use tools in a highly dexterous manner, without facilitation). Furthermore, there are striking similarities between Hawaiian and New Caledonian crows in terms of craniofacial morphology (Rutz et al., 2016) and prolonged juvenile dependency (Banko, Ball, & Banko, 2002). Importantly, New Caledonian and Hawaiian crows are not closely related (Jønsson, Fabre, & Irestedt, 2012), suggesting that their tool behaviour arose convergently, in response to similar ecological conditions prevailing on their tropical home islands (Rutz et al., 2016).

High-throughput genome sequencing has greatly aided the identification of genomic regions under selection and has thus improved our understanding of the molecular basis of adaptive evolution. Several approaches have been developed to quantify selection pressures acting on the genome. One of

the best-known methods is to test for deviations from neutral evolution in protein-coding regions, using the ratio of non-synonymous to synonymous substitutions – the dN/dS ratio where a ratio larger than one indicates positive selection and a ratio smaller than one, purifying selection (Kimura, 1977; Yang & Bielawski, 2000). This approach has been used widely to examine gene family evolution in several lineages, including *Drosophila* (Hahn, Han, & Han, 2007), mammals (Agaba et al., 2016; Park et al., 2015) and birds (Almén et al., 2016; Sackton et al., 2019). Strong positive selection acting on a locus will also leave a signature in neighbouring genomic regions. When a beneficial mutation increases in frequency and becomes fixed in the population, genetic variation in the vicinity of the advantageous mutation can be reduced or eliminated in a ‘selective sweep’ (Smith & Haigh, 1974). Selective sweeps have been detected in the patterns of genetic variation in the vicinity of genes associated with pigmentation, olfaction and immunity in humans (Williamson et al., 2007). Similarly, great tit (*Parus major*) genomes show evidence for selective sweeps in regions containing genes related to cognition and neuronal function (with one such gene (ERG1) also exhibiting a high dN/dS ratio indicative of positive selection; Laine et al., 2016), and selective sweeps are also implicated in the evolution of cognitive abilities in wasps (Miller et al., 2020). However, dN/dS may not be suitable for detecting selection in recently diverged species where sufficient numbers of substitutions may not have accumulated yet (Wolf, Künstner, Nam, Jakobsson, & Ellegren, 2009) and where polymorphism may bias its estimation (Mugal, Kutschera, Botero-Castro, Wolf, & Kaj, 2020). Moreover, the dN/dS ratio can only be estimated for protein-coding regions. In non-model organisms with limited ‘omics’ resources, genome scans for selective sweeps are a useful method to identify candidate genome regions for positive selection. Therefore, using both dN/dS and selective sweep detection approaches may better explore footprints of selection in closely related non-model organisms.

Here, we use a comparative genomics approach to search for genetic signatures of selection associated with traits facilitating tool use, by comparing the tool-using New Caledonian and Hawaiian crows to a sample of ten non-tool-using congeners. Specifically, we undertook *de novo* whole genome sequencing and assembly of *C. moneduloides* and whole-genome resequencing of multiple individuals from our set of *Corvus* species to: (1) establish a new genomically-informed phylogeny; (2) explore the phylogenetic relationship between tool-using and non-tool-using crows; (3) identify signatures of positive selection in New Caledonian and Hawaiian crows using phylogenetic and population genetic

approaches, as well as possible signatures of convergent molecular evolution; and (4) contrast the population genomics of tool-using New Caledonian crows with those of a closely-related and ecologically-matched non-tool-using species, the white-billed crow (*C. woodfordi*), which is endemic to the tropical Solomon Islands, north of New Caledonia. Our findings suggest that relatively subtle genetic changes can contribute to the evolution of traits facilitating tool use, including highly unusual craniofacial morphology.

## Materials and Methods

### Sampling and genome sequencing

We collated and generated data for a total of 12 crow species (see Supplemental Information; Tables S1-S2). Short read data are deposited at ENA (<https://www.ebi.ac.uk/ena>; ENA project accession numbers PRJEB33755 (umbrella project), PRJEB33706 and PRJEB33707), and metadata are available in Table S2.

### Reference genomes and annotation

We used a published reference assembly and its annotation for *C. (corone) cornix* (NCBI assembly accession GCF\_000738735.1; scaffold N50=16.4 Mb; Poelstra et al., 2014). For the purpose of this study, we also generated a *de novo* genome assembly of *C. moneduloides* (NCBI assembly accession number: GCA\_009650515.1) using ALLPATHS v47402 (Butler et al., 2008) as described in the Supplemental Information. We then annotated this assembly using the Maker2 v2.31.7 annotation pipeline (Holt & Yandell, 2011), using a dataset of 1,300 manually curated *C. (c.) cornix* proteins, a transcriptome dataset for *C. woodfordi* (<https://www.ebi.ac.uk/ena>; ENA project accession numbers PRJEB33755 (umbrella project), PRJNA577979) and *ab initio* gene predictions. The annotation is available on Dryad (DOI: <https://doi.org/10.5061/dryad.w0vt4b8m9>).

We assessed the quality of the assembly and annotation of the two reference genomes using the BUSCO v. 1.22 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) pipeline and the ‘vertebrata’ dataset (631 genes). We also identified orthologous coding sequences of the two annotations using orthAgogue v1.0.3 (Ekseth, Kuiper, & Mironov, 2014). Specifically, we identified

orthologs, paralogs and co-orthologs as described in Ekseth et al. (2014) (See Supplementary Material).

### **Short reads processing and protein-coding gene sets generation**

We trimmed adapters and mapped the newly sequenced and available short reads for the 12 crow species to both the *C. (c.) cornix* and *C. moneduloides* assemblies. We marked duplicates and realigned reads around indels (see Supplemental Information). Next, we generated consensus genomes for each of the 12 species from the resulting bam files, and following Mugal et al. (2020), masked polymorphic sites with 'Ns', thereby only considering fixed sites for tests of positive selection (PAML, McDonald Kreitman tests). For all downstream analyses, we only retained protein-coding sequences with at least 80% completeness (i.e., containing no more than 20% 'N' relative to the gene in the respective annotation) in order to avoid potential bias when calculating indices of selection. Because of varying data quality among re-sequenced genomes, not all consensus genomes had the same number of genes with  $\geq 80\%$  completeness. Therefore, in order to maximise the number of genes with  $\geq 80\%$  completeness, while also maintaining adequate sample sizes and statistical power, we generated datasets for 5, 7 and 8 species (see Supplemental Information).

### **Phylogenetic tree reconstruction**

We extracted sequences of 631 core vertebrate genes present in all species ('vertebrata' dataset from BUSCO; Simão et al., 2015) and aligned them in PRANK v100802 (Loytynoja & Goldman, 2005). We then used RAxML (Stamatakis, 2014) to perform a maximum likelihood analysis, assuming a GTR + gamma model of sequence evolution including four rate categories for each codon position and using 1,000 bootstrap iterations. *Taeniopygia guttata* was specified as the outgroup (see Supplemental Information).

### **Tests of positive selection**

We used two approaches to identify signatures of positive selection between habitually tool-using and non-tool-using species. First, we used codeml from the PAML package v4.9 (Yang, 2007) to test for different rates of evolution between crow lineages (Supplemental Information, Table S3). PAML

estimates dN/dS ( $\omega$ ), which compares the rate of substitutions at synonymous sites (dS), which are presumed neutral, to the rate of substitutions at non-synonymous sites (dN), which are under selection. A dN/dS > 1 is thus taken to imply the action of positive selection on a protein-coding gene, whereas a dN/dS < 1 indicates purifying selection. We used a branch model and considered a ‘null model’ assuming one  $\omega$  value across the entire tree, and an ‘alternative model’ assuming one separate  $\omega$  for some branches and another  $\omega$  for the remainder of the tree. Null and alternative models were compared using a Likelihood Ratio Test (LRT) with 1 degree of freedom. False Discovery Rate (FDR) for multiple testing corrections were applied (Benjamini, Heller, & Yekutieli, 2009), using the ‘qvalue’ R package (Storey, Bass, Dabney, & Robinson, 2019) at a FDR threshold of 0.05. We tested the hypothesis that a sequence evolved at different rates in the tool-using *C. moneduloides* or *C. hawaiiensis* lineages compared to the rest of our *Corvus* phylogeny. We ran these tests labelling only the *C. moneduloides* branch or labelling both *C. moneduloides* and *C. hawaiiensis* to have a different  $\omega$  from the rest of the tree (Table S3). For the latter model, we tested (1) a different  $\omega$  for each species and (2) the same  $\omega$  for both species, thereby testing for a signal of convergent positive selection in the two species (Yang & Bielawski, 2000; Yang, 2007; see Supplemental Information). In order to assess the effects of the varying quality of the assemblies and their annotation on our results, these tests were run twice, using genes extracted either from the *C. moneduloides* reference genome or from the *C. (c.) cornix* reference genome. This approach also enabled us to look for evidence of genetic novelties arising in *C. moneduloides* versus *C. (c.) cornix* that might correlate with tool use. For each of the two reference genome assemblies, we compiled three different species datasets that used comparisons among 5, 7, or 8 species of the genus *Corvus* together with an outgroup, the zebra finch (*T. guttata*). Compiling these three subset of our data allowed us to (1) maximise the number of genes with 80% completeness and maintain adequate sample sizes and statistical power (Table S3; see above and Supplemental Information); and (2) minimise the discovery of false positives associated with ancestral and lineage-specific polymorphisms (Mugal et al., 2020; Mugal, Wolf, & Kaj, 2014) by excluding *C. woodfordi*, the closest related species of *C. moneduloides*, from these analyses (Fig. 1, S1). To ensure that analyses using different reference genomes produced similar results, we compared  $\omega$  values obtained under the null model from analyses using different reference genomes for individual genes. We also compared the average  $\omega$  of genes within Gene Ontology (GO) terms (see



Supplemental Information). Finally, we compared  $\omega$  values obtained under the alternative model using different reference genomes for the *C. moneduloides* lineage and for the background lineages.

As a second approach to identify positive selection, we applied the McDonald-Kreitman framework (McDonald & Kreitman, 1991; Mugal et al., 2014) to test for evidence of selection in *C. moneduloides*, contrasting divergence with polymorphism data, based on the *C. moneduloides* genome assembly and annotation. Variant discovery and genotyping were performed using the GATK pipeline v3.4.0 (DePristo et al., 2011; Van der Auwera et al., 2013) from processed bam files as described in the Supplemental Information. The VCF file was annotated using snpEff v4.2 (Cingolani et al., 2012). Counts of synonymous (pS) and non-synonymous (pN) polymorphic changes for each gene were extracted from the VCF file and merged with counts of synonymous (dS) and non-synonymous (dN) substitutions for the branch leading to *C. moneduloides* in different PAML runs (5, 7, or 8 species; as described above). We tested for significant deviations from neutrality for each gene by performing Fisher's exact tests in R v3.4.2 (R Core Team, 2019). FDR corrections were applied with a threshold of 0.05. Further, we calculated the neutrality index  $NI=(pN/dN)/(pS/dS)$ . In the case of neutral evolution, the ratio of nonsynonymous to synonymous substitutions and polymorphisms should be represented at equal proportions resulting in  $NI=1$ . Conversely, for genes under positive selection and indicating an excess of non-synonymous fixed differences, this index will result in a value of  $NI<1$  (Rand & Kann, 1996). For further details, see Supplemental Information.

### **Pathway and biological function analysis.**

To examine the functional implications of our results, we explored Gene Ontology (GO) and performed a statistical overrepresentation test of biological processes in PANTHER (Mi, Muruganujan, Casagrande, & Thomas, 2013). We used as input all candidate genes indicative of positive selection identified with PAML and McDonald-Kreitman tests in *C. moneduloides* or *C. hawaiiensis*. We used Fisher's exact test with multiple FDR correction and a FDR threshold of 0.05.

### **Population genomic analyses in *C. moneduloides* and *C. woodfordi***

*C. moneduloides* and *C. woodfordi* are closely related and inhabit ecologically similar tropical islands, but only the former is known to use tools for extractive foraging (Jønsson et al., 2012). A comparison

of genome-wide diversity in coding and non-coding regions within these species, which experience comparable levels of mutational load (Kutschera et al., 2020), may thus reveal evidence of more recent selection that is unique to *C. moneduloides*. Selective sweeps should leave a signature of locally reduced diversity. Such reductions in diversity that are unique to *C. moneduloides* (and do not appear in *C. woodfordi*) would point to selection pressures other than those associated with living on tropical islands, possibly including selection on tool-use-related traits. Therefore, we examined low-diversity genomic regions from processed bam files using ANGSD v0.917 (Korneliussen, Albrechtsen, & Nielsen, 2014) and calculated genome-wide diversity indices for *C. moneduloides* and *C. woodfordi* (see Supplemental Information). To polarise the site frequency spectrum (SFS), we reconstructed an ‘ancestral’ consensus genome using all the available data from the *C. corone* species group, *C. frugilegus*, and *C. dauuricus*. This ancestral consensus genome was assembled as for the other species (see Supplemental Information). We then estimated Tajima’s *D*, and Fay and Wu’s *H* in overlapping windows of 50kb with a step size of 10kb based on this polarised SFS. We used a threshold for Tajima’s *D* of -1.51 based on population genetic simulations (see Supplemental Information). To check for evidence of population structure, we ran a PCA with PLINK v1.9 (Chang et al., 2015).

## Results

### Comparison of reference genome assemblies and annotations

The *de novo* assembled genome of the New Caledonian crow (*C. moneduloides*) had a total size of 1.01 Gb, a contig N50 of 331 kb and a scaffold N50 of 4.5 Mb (SI Materials and Methods). Overall, the reference genome annotations contained 14,149 and 18,649 protein-coding gene models for *C. (c.) cornix* and *C. moneduloides*, respectively. Mean gene model length is 1,416 bp for *C. moneduloides* and 1,759 bp for *C. (c.) cornix*. We examined homology between protein-coding sequences from the two reference genomes with orthAgo (Ekseth et al., 2014) resulting in 12,100 orthologs. Out of these, 8,489, 3,714 and 1,942 were unique orthologs, paralogs and co-orthologs, respectively (see Supplemental Information).

### Phylogenetic relationships

The topology of our phylogenetic tree based on 631 core vertebrate genes was broadly consistent with earlier phylogenetic reconstructions based on fewer molecular markers (Fig. 1,S1; Haring et al., 2012; Jönsson et al., 2012; Rutz et al., 2016). However, there are some notable differences; for example, in our phylogeny *C. hawaiiensis* groups with *C. corax*, rather than with *C. frugilegus* (Fig. 1,S1).

### Phylogeny-based tests of positive selection

Next, we screened for genes with evidence of positive selection using comparative genomic analyses. We mapped re-sequencing data from 12 *Corvus* species to either the *C. (c.) cornix* or the *C. moneduloides* reference genome and extracted all protein-coding sequences. Using the *C. moneduloides* reference genome, for which we generated an annotation of 18,649 protein-coding genes, we recovered 17,967, 17,487 and 17,155 genes with  $\geq 80\%$  completeness for the 5-, 7- and 8-species datasets, respectively. With the *C. (c.) cornix* reference genome, we recovered 10,138, 10,096, and 10,070 genes with  $\geq 80\%$  completeness for the 5-, 7-, and 8-species datasets, respectively (Table S3).

Using these datasets and branch models of codon substitution rates, we tested for evidence of positive selection on branches leading to *C. moneduloides* and *C. hawaiiensis* as well as for convergent evolution between them. Overall, we found evidence of positive selection in *C. moneduloides* in 12 genes: five genes when using *C. moneduloides* as the reference genome and seven genes when using *C. (c.) cornix* as the reference genome (Table 1). Among the five genes identified using the *C. moneduloides* reference genome, we found two genes (CAPN15, CAMLG), one gene (TPBGL) and two genes (SPSB1 and CIITA) under putative positive selection ( $\omega > 1$ ) in the 5-, 7- and 8-species datasets, respectively (Table 1). We found only one gene (CIITA) under putative positive selection in *C. moneduloides* and *C. hawaiiensis* when assuming the same substitution rate for both species (Table 1).

Among the seven genes identified using the *C. (c.) cornix* reference genome, we found three (LHX1, ZC2HC1A, and DTNBP1), two (HTRA2, TBC1D10A) and two genes (NPRL2, SLC6A20) with  $\omega > 1$  as candidates for positive selection in *C. moneduloides* using the 5-species, 7- and 8-species dataset, respectively (Table 1). It is worth noting that several other genes have unreasonably high  $\omega$  values, indicative of the low number of synonymous changes in crows reducing the

denominator in dN/dS to near zero (Table 1), which can occur when comparing closely related species (Wolf et al., 2009).

While candidate genes overlapped between species sets, no candidate genes were shared among the analyses using the two different reference genomes. This lack of overlap may result from a genuine absence of these genes due to varying quality of the respective annotations, or it may simply be explained by the fact that there were only 12,100 orthologs in common between the two annotations. Additionally, the lack of overlap may be caused by incomplete lengths of gene models in both genomes' annotations. We therefore tested whether there was any correlation in the estimates of  $\omega$  under the null model (a single common  $\omega$  for the entire tree) between analyses using two different reference genomes. Correlations were strong for all species datasets (Fig. S2). Strong correlations were also evident when averaging  $\omega$  values of individual genes across GO terms (Spearman rank correlations between 0.80 and 0.81 across all datasets; Fig. S2). In contrast, the correlations for  $p$ - and  $q$ -values of individual genes were much weaker (Spearman rank correlations between 0.20 and 0.57 across all datasets). Correlations of  $\omega$  estimates for the *C. moneduloides* lineage in alternative models across analyses using different reference genomes were also much lower ( $\sim 0.5$ ; Fig. S3). Moreover, in datasets using the *C. (c.) cornix* reference genome, pairwise correlations of  $\omega$  estimates for the *C. moneduloides* lineage from three separate software runs were high but not perfect ( $\sim 0.8$ ), while correlations of  $\omega$  estimates for the background lineages were much better ( $\sim 0.9$ ; Table S4).

### McDonald-Kreitman tests

We contrasted between-species substitutions on the *C. moneduloides* branch obtained from the PAML analyses with within-species polymorphism based on the *C. moneduloides* reference genome. Across three different datasets, we found a total of 14 candidate genes with NI smaller than one (Rand & Kann 1996) at  $p < 0.05$  (Fisher's exact test) (Table S5). Of these, 11 genes were present in the 5-species dataset (17,967 genes) and one gene (LOC100232337) was present in the 7-species dataset (17,487 genes). This same gene also formed part of the results from the 8-species dataset (17,096 genes) based on the model with *C. moneduloides* and *C. hawaiiensis* with different substitution rates, along with two more genes (NCOR2, PTPN9) with  $NI < 1$  and  $p < 0.05$  (Fisher's exact test) (Table S5). After applying the false discovery rate (FDR) correction for multiple testing ( $q = 0.05$ ), no gene with

NI<1 remained significant for any of the datasets. None of the 14 candidate genes identified in the McDonald-Kreitman test showed a signal for positive selection in the phylogeny-based analyses using PAML.

### **GO term over-representation analysis**

Over-representation analysis based on gene function (gene ontology biological processes) showed that genes with elevated  $\omega$  (mostly in *C. moneduloides*; Table 1) and with an excess of non-synonymous fixed differences (McDonald-Kreitman tests; Table S5) were associated with cell differentiation, growth and development, brain and nervous system, and retinal development (e.g., LHX1; Table S6).

### **Comparative population genomics in a tool-using and a non-tool-using crow species**

We calculated population genetic diversity statistics for *C. moneduloides* and *C. woodfordi* in 50 kb sliding windows with 10 kb overlap. Overall, genetic diversity ( $\pi$ ) is lower in *C. woodfordi* (mean and 95% CI, Autosomes: 0.00036 [0.000356, 0.000359], Z-Chromosome: 0.00026 [0.000256, 0.000268]) than in *C. moneduloides* (A: 0.0011 [0.00105, 0.00106], Z: 0.00081 [0.000803, 0.000822]). The relationship between genetic diversity and chromosome length is strongly negative in both species (*C. moneduloides*:  $\rho=-0.73$ ,  $p < 0.001$ ; *C. woodfordi*:  $\rho=-0.9$ ,  $p<0.001$ ). As in other avian lineages (Dutoit et al., 2017; Ellegren, 2013; Weissensteiner et al., 2017), larger chromosomes have lower diversity, on average, than smaller chromosomes. At the same time, Tajima's  $D$  was on average slightly lower in *C. woodfordi* (Autosomes: 0.180 [0.175, 0.184], Z chr.: -0.046 [-0.0623, -0.0341]) than in *C. moneduloides* (Autosomes: 0.182 [0.179, 0.185], Z chr.: 0.32 [0.304, 0.334]). In both species, mean Tajima's  $D$  was consistently greater than zero (Fig. 2). Fay and Wu's  $H$  was on average higher than zero in *C. moneduloides* (except on the Z-chromosome), while in *C. woodfordi* values were consistently lower than zero (Fig. S4).

In *C. moneduloides*, we found 81 genomic windows with Tajima's  $D$  lower than what would be expected for neutrally evolving regions based on population genetic simulations (see Methods; Fig. S5). PCA analysis did not reveal any obvious patterns of population structure in *C. moneduloides* (Fig. S6A) despite individual samples originating from different islands and different sampling years. In *C. woodfordi*, samples were differentiated along the first PC axis according to sampling location

(Fig. S6B). Undetected population admixture would increase values of Tajima's  $D$  (Huber & Lohmueller, 2016; Tajima, 1993), although this would make the simulation-based threshold conservative. Extracting all candidate windows with Tajima's  $D < -1.51$  in *C. moneduloides* as potentially indicating a selective sweep, as well as 1 Mb up- and down-stream, and then merging all overlapping windows into contiguous regions, we obtained 11 wider regions spanning an average of ~2Mb (Fig. 2b). In several of these windows, Tajima's  $D$  was only reduced in *C. moneduloides* (highlighted in grey in Fig. 2b), and in these same regions Fay and Wu's  $H$  was also markedly reduced only in *C. moneduloides*. The 11 wider regions contained a total of 350 genes (Table S7), among which three genes have previously been associated with craniofacial or bill morphology evolution in avian lineages (CALM1, PPP4R4, DKK2; Abzhanov et al., 2006) and two genes with the mouse 'small maxilla' GO term (TRPS1, Suemoto et al., 2007; CHST11; Klüppel, Wight, Chan, Hinek, & Wrana, 2005; Fig. 2b). Over-representation analysis (GO) of biological processes of these 350 genes revealed an over-representation of genes related to glycosylation (Table S8). However, none of these 350 genes showed evidence of positive, divergent selection in our PAML analyses or McDonald-Kreitman tests (see above).

## Discussion

There are over 40 species of crows and ravens worldwide (Haring et al., 2012). While this taxonomic group exhibits surprisingly conserved morphology and behaviour, habitual tool use seems to have evolved in only two island species: the New Caledonian and the Hawaiian crow (Hunt, 1996; Rutz et al. 2016). Here, we sequenced the genomes of 12 crow species, established a phylogeny from these data, and used comparative genomic approaches to search for genomic signatures of a tool-using lifestyle in coding sequences. While we detected signals of positive selection near genes associated with bill morphology (as expected from earlier work; Matsui et al., 2016; Troscianko et al., 2012), genetic changes associated with tool behaviour in crows appear subtle. Perhaps, key adaptations do not require extensive changes in protein-coding regions but are rather underpinned by several genes of small effect or differences in gene regulation. Finally, the lack of concordance among our analyses using different reference genomes, which vary in quality for mapping re-sequencing data from different species, illustrates the limitations of using a single reference assembly for tests of positive

selection and offers an important cautionary note to interpretations made using these unidirectional comparisons.

Phylogenetic reconstruction based on substitutions in 631 genes confirmed that *C. moneduloides* and *C. hawaiiensis* are distantly related, suggesting that their tool behaviour evolved convergently (Rutz et al., 2016). The interspecific relationships identified within the genus differ slightly compared to earlier phylogenies based on a few nuclear and mitochondrial markers (Haring, Gamauf, & Kryukov, 2007; Jönsson et al., 2012).

Across all species comparisons, we identified candidate signals of positive selection in 12 genes associated with cell differentiation, growth and development, and involved in brain, nervous system and eye development. Positive selection was identified mostly in *C. moneduloides*, while only one gene coding for class II major histocompatibility complex transactivator (CIITA), showed weak but convergent signals for positive selection in *C. moneduloides* and *C. hawaiiensis*. We further investigated population genomic signals of selection in *C. moneduloides* and *C. woodfordi*, an ecologically-matched non-tool-using sister species. These analyses identified genomic regions with low genetic diversity (i.e., low Tajima's  $D$  and Fay and Wu's  $H$ ) in *C. moneduloides*, proximal to genes associated with bill morphology, potentially as a result of recent selection. This could be indicative of selective sweeps (Fay & Wu, 2000; Kim & Stephan, 2002; Smith & Haigh, 1974; Stephan, 2016) that occurred as a result of selection on genes underlying variation in bill morphology that enhances tool manipulation skills (Matsui et al., 2016; Troscianko et al., 2012). The overall positive values of Tajima's  $D$  and negative Fay and Wu's  $H$  on a genome-wide level suggest that we may have sampled structured populations in both species with unequal representation (Huber & Lohmueller, 2016; Tajima, 1993). In fact, an earlier study has found significant local structuring of *C. moneduloides* populations (Rutz, Ryder, & Fleischer, 2012), although island-scale patterns appeared weaker (Abdelkrim, Hunt, Gray, & Gemmell, 2012). No data on population structure are available for *C. woodfordi*, but significant structuring has recently been reported for another crow species inhabiting a remote tropical island, the Mariana crow (*C. kubaryi*; Cortes-Rodriguez et al., 2019). Curiously, Tajima's  $D$  is negative on the  $Z$  chromosome in *C. woodfordi* which might be due to sex-biased dispersal or strong purifying selection. A biased sampling of heterogametic females bearing on diversity estimates is not a likely cause for this negative Tajima's  $D$  since four out of five *C.*

*woodfordi* were males. A complete description of population structure and gene-flow in these species will require more systematic sampling of more individuals, which was outside the scope of the current study.

The subtle genetic changes identified for tool-using New Caledonian and Hawaiian crows are consistent with the observation that crows in general have large brains compared to other birds when corrected for body size, and exhibit notable behavioural plasticity and innovation potential (Emery & Clayton 2004). As noted in the introduction, some crow species that do not routinely use tools in the wild, such as rooks (*C. frugilegus*), can use stick-type tools under laboratory conditions, usually following some facilitation (e.g., Bird & Emery, 2009). Thus, the cognitive ability for basic tool use probably arose in the common ancestor of *Corvus* (Lefebvre, Nicolakakis, & Boire, 2002). We found evidence for positive selection and selective sweeps near genes that control bill morphology and development, but not near genes associated with brain function. Bill shape is one of the most striking differences between habitually tool-using and non-tool-using crows (Matsui et al., 2016; Troscianko et al., 2012; Rutz et al. 2016; Fig.1). Several lines of evidence suggest that straight bills and highly-mobile (and large) eyes facilitate the handling of stick-type tools (Matsui et al., 2016; Troscianko et al., 2012). Indeed, non-tool-using rooks are not as dexterous as *C. moneduloides* and *C. hawaiiensis* at using tools, presumably partly because they lack enabling bill features (Bird & Emery, 2009). At present, it is unclear if selection for specialized craniofacial features preceded (e.g., to enable woodpecker-like chiselling of deadwood; for Rutz & St Clair, 2012), and/or followed (Hunt & Gray, 2007; Matsui et al., 2016), the emergence of tool behaviour (Troscianko et al., 2012). In any case, while morphological change seems a necessary component for the early evolution of habitual dexterous tool behaviour, more subtle changes affecting ontogenetic programmes and cognitive capacities may also occur, but remained undetected in our present analyses.

There is high conservation in gene sequence and chromosomal synteny in birds (Zhang et al., 2014), and the crow lineage diverged relatively recently (~17.5 mya), with the last common ancestor of *C. moneduloides* and *C. hawaiiensis* dated to the mid-Miocene (~11 mya; Jönsson et al., 2012). It seems probable, therefore, that the relatively short evolutionary time frames since their divergence may have limited the possibility for repeated changes in coding regions, and thus resulted in the low number of overall substitutions among the crow lineages, leading to low statistical power in some of



our tests. This is consistent with the lack of a strong signal for positive selection we observed using dN/dS analyses and McDonald-Kreitman tests. Moreover, our results indicate that, while single estimates of  $\omega$  were robust to the reference genome under the null model, these estimates varied substantially according to the chosen reference genome for the alternative models. This may further explain the lack of overlap in the identified candidate genes from our analyses using two different reference genomes and further indicates that our dN/dS analyses may be limited in resolution and power. Importantly, the lack of overlap in candidate genes between dN/dS analyses illustrates the limitations associated with using a single assembly for tests of positive selection. Because using one or the other assembly led to inconsistent results, different interpretations and conclusions on the link between genomic changes and a trait could have been drawn. However, mapping re-sequenced data to two different reference genomes and performing reciprocal dN/dS analyses enabled us to determine whether the signals of positive selection in coding genes observed have a robust biological basis or instead result from mapping or analysis bias. We thus caution against the use of a single reference genome and strongly recommend that reciprocal tests be performed whenever possible, especially when several assemblies of varying qualities (i.e., contiguity, coverage) are available. Ideally, reference genomes should be assembled for each species under consideration along with population genomic data to also account for within species variation.

It is possible that genomic changes underlying recent adaptations across this phenotypically conserved group of birds may not be reflected in extensive coding sequence divergence but instead in regulatory regions (Mack & Nachman, 2017; Wittkopp & Kalay, 2011). This, too, is consistent with our findings of signatures of selective sweeps up- and downstream of candidate genes and a mixed signal within coding sequences, although this interpretation should be treated cautiously due to low sample sizes. Analysis of regulatory motifs, such as conserved non-exonic elements (CNEEs), would enable more refined examination of the genomic adaptations to tool behaviour in crows. However, the need for large sample sizes, high-quality assemblies extending to non-coding elements for all crow species, as well as high-quality gene annotations, means a robust analysis of such CNEEs is beyond the scope of the present study.

The importance of regulatory changes in driving phenotypic differences (King & Wilson, 1975) has become increasingly obvious as relevant genomic data have become available. Work in

humans, rodents, flies and fish indicate that regulatory variants have broad influence on the transcriptome and the proteome, sexual dimorphism, skin colour, behaviour, disease susceptibility, and cognition (Albert & Kruglyak, 2015; Enoch et al., 1998; Mank, 2017; Miller et al., 2007; Miller et al., 2020; Wray, 2007). A recent study examining the loss of flight in ratites, for example, adds to a growing body of evidence suggesting that changes in regulatory regions are frequently more important than changes in protein-coding genes (Sackton et al., 2019). While coding sequence variation is important in some classes of genetic divergence (e.g., immune function; Blais et al., 2007), and the relative importance of coding and non-coding divergence has been debated, regulatory changes may be of primary importance for phenotype evolution in very recently diverged species (Mack & Nachman, 2017; Wittkopp & Kalay, 2011).

One challenge for studies on the genetic basis of complex traits, such as tool behaviour, which may not only be associated with specialised morphology but also with developmental, cognitive and other behavioural adaptations, is that such traits are likely polygenic and not underpinned by strong divergent selection on only one or a few genes (Nagel et al., 2018; Savage et al., 2018). This implies that changes in the frequency of advantageous alleles may be subtle and thus hard to detect (Pritchard, Pickrell, & Coop, 2010; Stephan, 2016). Moreover, epigenetic processes are known to regulate genes associated with a wide range of traits, including some associated with cognition (Goldberg, Allis, & Bernstein, 2007). Methylation patterns, one form of epigenetic regulation, have been shown to correlate with higher expression of neuronal genes in humans (Lister et al., 2009). Great tits (*Parus major*), for example, a species that is a useful model for cognitive studies (Laine et al., 2016) have reduced CpG and non-CpG methylation signatures similar to those found in the neuronal tissues of mammals (Guo et al., 2014; Laine et al., 2016). Consequently, further investigation of the molecular basis of adaptations associated with the tool-using and -making abilities of New Caledonian and Hawaiian crows would likely benefit from the examination of both genetic and epigenetic changes in coding and regulatory regions.

Taken together, our results indicate that the genetic changes associated with habitual tool-related behaviour in crows are surprisingly subtle, perhaps reflecting, in part, the difficulty in identifying signatures of selection in recently diverged lineages, and the fact that crows, in general, share a common capacity for behavioural innovation. However, at least one of our findings is

consistent with a potential adaptation to tool behaviour in New Caledonian and Hawaiian crows - an unusually straight bill. Investigating the genetic basis of traits that can facilitate seemingly sophisticated skills such as advanced tool behaviour is challenging. In fact, traits such as the unusual bill shape of New Caledonian crows and (to a lesser extent) Hawaiian crows are very likely to be underpinned by numerous genes of small effect and not by a few genes of large effect, as has been documented for speech (Atkinson et al., 2018; Krause et al., 2007), intelligence and personality (Nagel et al., 2018; Savage et al., 2018). Furthermore, with the realisation that changes in regulatory regions can be associated with developmental pathways (Sackton et al., 2019), the exploration of genetic and epigenetic changes in regulatory regions will be instrumental in understanding the evolution and genetic basis of a variety of traits that may not be obvious based on the examination of protein-coding regions alone.

### **Acknowledgements**

A David Phillips Fellowship to C.R. from the UK's Biotechnology and Biological Sciences Research Council (BBSRC; grant BB/G023913/2) provided funding for the sequencing of one of the *C. moneduloides* genomes (FS66096), and the genomes of *C. hawaiiensis*, *C. kubaryi*, *C. tasmanicus*, and *C. corax*. We thank Nancy McInerney of the Smithsonian Center for Conservation Genomics for preparing Illumina libraries of the latter four taxa. Further funding for personnel and data generation of the remaining species (except *C. moneduloides* and *C. woodfordi*) was provided by the European Research Council (ERCStG-336536 FuncSpecGen to J.B.W.W.), the Swedish Research Council Vetenskapsrådet (621-2013-4510 to J.B.W.W.), the Knut and Alice Wallenberg Foundation (to J.B.W.W.), the Lawski foundation (to V.E.K. and J.B.W.W.) and the German Research Foundation (KU 3402/1-1 to V.E.K.). Analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). A Marsden Fund Grant to G.R.H., R.D.G. and N.J.G. from the Royal Society of New Zealand (UOA1208), a Japanese Society for Promotion of Science Postdoctoral Fellowship (H.A.), together with funding from University of Auckland (G.R.H. and R.D.G.), the Department of Linguistic and Cultural Evolution at the Max Planck Institute for the Science of Human History, and University of Otago (N.J.G.), supported personnel, in the field sampling, and the sequencing and *de novo* assembly of the *C. moneduloides* genome, the resequencing of *C. moneduloides* and *C. woodfordi* genomes for population analyses, and the generation of transcriptome resources for *C. woodfordi*. N.D. acknowledges funding from the Swiss National Science Foundation (P2SKP3\_165031 and P300PA\_177845) and the Carl Tryggers Foundation

(CTS 19:257). R.A.W.W. acknowledges combined Natural Environment Research Council and St Andrews 600th Anniversary PhD Studentship grants (NE/L501852/1).

## References

- Abdelkrim, J., Hunt, G. R., Gray, R. D., & Gemmell, N. J. (2012). Population genetic structure and colonisation history of the tool-using New Caledonian crow. *PloS One*, *7*(5), e36608.
- Abzhanov, A., Kuo, W. P., Hartmann, C., Rosemary Grant, B., Grant, P. R., & Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, *442*(7102), 563-567. doi: 10.1038/nature04843
- Agaba, M., Ishengoma, E., Miller, W. C., McGrath, B. C., Hudson, C. N., Bedoya Reina, O. C., ... Cavener, D. R. (2016). Giraffe genome sequence reveals clues to its unique morphology and physiology. *Nature Communications*, *7*(1), 1-8.
- Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews. Genetics*, *16*(4), 197–212.
- Almén, M. S., Lamichhaney, S., Berglund, J., Rosemary Grant, B., Grant, P. R., Webster, M. T., & Andersson, L. (2016). Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *BioEssays*, *38*(1), 14-20. doi: 10.1002/bies.201500079
- Atkinson, E. G., Audesse, A. J., Palacios, J. A., Bobo, D. M., Webb, A. E., Ramachandran, S., & Henn, B. M. (2018). No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. *Cell*, *174*(6), 1424–1435.e15.
- Banko, P. C., Ball, D. L., & Banko, W. E. (2002). Hawaiian Crow (*Corvus hawaiiensis*). *The Birds of North America Online*. doi: 10.2173/bna.648
- Benjamini, Y., Heller, R., & Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1906), 4255-4271. doi: 10.1098/rsta.2009.0127
- Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370–10375.
- Biro, D., Haslam, M., & Rutz, C. (2013). Tool use as adaptation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1630), 20120408.
- Blais, J., Rico, C., van Oosterhout, C., Cable, J., Turner, G. F., & Bernatchez, L. (2007). MHC adaptive

- divergence between closely related and sympatric African cichlids. *PloS One*, 2(8), e734.
- Bluff, L. A., Troscianko, J., Weir, A. A., Kacelnik, A., & Rutz, C. (2010). Tool use by wild New Caledonian crows *Corvus moneduloides* at natural foraging sites. *Proceedings of the Royal Society B: Biological Sciences*, 277(1686), 1377-1385.
- Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I. A., Belmonte, M. K., Lander, E. S., ... Jaffe, D. B. (2008). ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Research*, 18(5), 810–820.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.
- Cnotka, J., Güntürkün, O., Rehkämper, G., Gray, R. D., & Hunt, G. R. (2008). Extraordinary large brains in tool-using New Caledonian crows (*Corvus moneduloides*). *Neuroscience Letters*, 433(3), 241–245.
- Cortes-Rodriguez, N., Campana, M., Berry, L., Faegre, S., Derrickson, S., Ha, R., ... Fleischer, R. (2019). Population Genomics and Structure of the Critically Endangered Mariana Crow (*Corvus kubaryi*). *Genes*, 10(3), 187. doi: 10.3390/genes10030187
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498.
- Dutoit, L., Vijay, N., Mugal, C. F., Bossu, C. M., Burri, R., Wolf, J., & Ellegren, H. (2017). Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proceedings. Biological Sciences / The Royal Society*, 284(1849). doi: 10.1098/rspb.2016.2756
- Ekseth, O. K., Kuiper, M., & Mironov, V. (2014). orthAgo: an agile tool for the rapid prediction of orthology relations. *Bioinformatics*, 30(5), 734–736.
- Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Annual Review of Ecology, Evolution, and Systematics*, 44, 239-259. doi: 10.1146/annurev-ecolsys-110411-160327
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science*, 306(5703), 1903-1907.
- Enoch, M. A., Kaye, W. H., Rotondo, A., Greenberg, B. D., Murphy, D. L., & Goldman, D. (1998). 5-HT2A promoter polymorphism -1438G/A, anorexia nervosa, and obsessive-compulsive disorder. *The Lancet*, 351(9118), 1785–1786.

- Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413.
- Gallot, Q., & Gruber, T. (2019). Spontaneous use and modification of a feather as a tool in a captive common raven. *Ethology*, *125*(10), 755-758
- Gibson, K. R., Gibson, K. R., & Ingold, T. (1994). *Tools, Language and Cognition in Human Evolution*. Cambridge University Press.
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell*, *128*(4), 635–638.
- Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R. D., ... Taylor, A. H. (2019). New Caledonian Crows Use Mental Representations to Solve Metatool Problems. *Current Biology: CB*, *29*(4), 686–692.e3.
- Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., ... Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, *17*(2), 215–222.
- Hahn, M. W., Han, M. V., & Han, S.-G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, *3*(11), e197.
- Haring, E., Däubel, B., Pinsker, W., Kryukov, A., & Gamauf, A. (2012). Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae) - a first survey based on museum specimens. *Journal of Zoological Systematics and Evolutionary Research*, *50*(3), 230-246.
- Haring, E., Gamauf, A., & Kryukov, A. (2007). Phylogeographic patterns in widespread corvid birds. *Molecular Phylogenetics and Evolution*, *45*(3), 840–862.
- Henrich, J. (2015). *The Secret of Our Success*. doi: 10.2307/j.ctvc77f0d
- Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491.
- Holzhaider, J. C., Sibley, M. D., Taylor, A. H., Singh, P. J., Gray, R. D., & Hunt, G. R. (2011). The social structure of New Caledonian crows. *Animal Behaviour*, *81*(1), 83-92. doi: 10.1016/j.anbehav.2010.09.015
- Huber, C. D., & Lohmueller, K. E. (2016). Population Genetic Tests of neutral Evolution. In *Encyclopedia of Evolutionary Biology* (pp. 112–118).
- Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows. *Nature*, *379*(6562), 249-251. doi: 10.1038/379249a0
- Hunt, G. R., & Gray, R. D. (2002). Species-wide manufacture of stick-type tools by New Caledonian Crows. *Emu - Austral Ornithology*, *102*(4), 349-353. doi: 10.1071/mu01056
- Hunt, G. R., & Gray, R. D. (2003). Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proceedings. Biological Sciences / The Royal Society*, *270*(1517), 867–874.

- Hunt, Gavin R., and Russell D. Gray. (2007). Genetic assimilation of behaviour does not eliminate learning and innovation. *Behavioral and Brain Sciences* 30(4), 412-413.
- Hunt, G. R., Gray, R. D., & Taylor, A. H. (2010). Why is tool use rare in animals? *Tool Use in Animals*, pp. 89–118. doi: 10.1017/cbo9780511894800.007
- Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2012). Prolonged Parental Feeding in Tool-Using New Caledonian Crows. *Ethology*, 118(5), 423-430. doi: 10.1111/j.1439-0310.2012.02027.x
- Hunt, G. R., & Uomini, N. (2016). A complex adaptive system may be essential for cumulative modifications in tool design. *Japanese Journal of Animal Psychology*, 66(2), 141-159.
- Jönsson, K. A., Fabre, P.-H., & Irestedt, M. (2012). Brains, tools, innovation and biogeography in crows and ravens. *BMC Evolutionary Biology*, 12, 72.
- Kanai, M., Matsui, H., Watanabe, S., & Izawa, E.-I. (2014). Involvement of vision in tool use in crow. *Neuroreport*, 25(13), 1064–1068.
- Kaplan H, Gurven M, and Winking J. (2009). An evolutionary theory of human lifespan: embodied capital and the human adaptive complex. In Bengtson V, Silverstein M, Putney N, and Gans D (Ed.), *Handbook of theories of aging, 2nd Edition* (pp. 39–66). Springer, NY.
- Kenward, B., Rutz, C., Weir, A. A. S., & Kacelnik, A. (2006). Development of tool use in New Caledonian crows: inherited action patterns and social influences. *Animal Behaviour*, 72(6), 1329-1343. doi: 10.1016/j.anbehav.2006.04.007
- Kenward, B., Weir, A. A. S., Rutz, C., & Kacelnik, A. (2005). Behavioural ecology: tool manufacture by naive juvenile crows. *Nature*, 433(7022), 121.
- Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, 267(5608), 275–276.
- Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2), 765–777.
- King, M., & Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. *Science*, 188(4184), 107-116. doi: 10.1126/science.1090005
- Klump, B. C., Masuda, B. M., St Clair, J. J. H., & Rutz, C. (2018). Preliminary observations of tool-processing behaviour in Hawaiian crows *Corvus hawaiiensis*. *Communicative & Integrative Biology*, 11(4), e1509637. doi: 10.1080/19420889.2018.1509637
- Klüppel, M., Wight, T. N., Chan, C., Hinek, A., & Wrana, J. L. (2005). Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. *Development*, 132(17), 3989–4003.

- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*, 356.
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., ... Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology*, *17*(21), 1908–1912.
- Kutschera, V. E., Poelstra, J. W., Botero-Castro, F., Dussex, N., Gemmell, N. J., Hunt, G. R., ... Wolf, J. B. W. (2020). Purifying Selection in Corvids Is Less Efficient on Islands. *Molecular Biology and Evolution*, *37*(2), 469–474.
- Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., ... Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*, 10474.
- Lefebvre, L., Nicolakakis, N., & Boire, D. (2002). TOOLS AND BRAINS IN BIRDS. *Behaviour*, *139*(7), 939–973. doi: 10.1163/156853902320387918
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., ... Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322.
- Loytynoja, A., & Goldman, N. (2005). From The Cover: An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences*, *102*(30), 10557–10562. doi: 10.1073/pnas.0409137102
- Mack, K. L., & Nachman, M. W. (2017). Gene Regulation and Speciation. *Trends in Genetics: TIG*, *33*(1), 68–80.
- Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nature Ecology & Evolution*, *1*(1), 6.
- Matsui, H., Hunt, G. R., Oberhofer, K., Ogihara, N., McGowan, K. J., Mithraratne, K., ... Izawa, E.-I. (2016). Adaptive bill morphology for enhanced tool manipulation in New Caledonian crows. *Scientific Reports*, *6*, 22776.
- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, *351*(6328), 652–654.
- Medina, F. (2013). *Study of the cognition and its neural substrate in New Caledonian crows* (PhD thesis). University of Auckland, New Zealand.
- Mehlhorn, J., Hunt, G. R., Gray, R. D., Rehkämper, G., & Güntürkün, O. (2010). Tool-making New Caledonian crows have large associative brain areas. *Brain, Behavior and Evolution*, *75*(1), 63–70.



- Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, 8(8), 1551–1566.
- Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D., & Kingsley, D. M. (2007). cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*, 131(6), 1179–1189.
- Miller, S. E., Legan, A. W., Henshaw, M. T., Ostevik, K. L., Samuk, K., Uy, F. M. K., & Sheehan, M. J. (2020). Evolutionary dynamics of recent selection on cognitive abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 117(6), 3045–3052.
- Mugal, C. F., Kutschera, V. E., Botero-Castro, F., Wolf, J. B. W., & Kaj, I. (2020). Polymorphism data assist estimation of the non-synonymous over synonymous fixation rate ratio  $\omega$  for closely related species. *Molecular Biology and Evolution*, 37(1), 260-279. doi: 10.1093/molbev/msz203
- Mugal, C. F., Wolf, J. B. W., & Kaj, I. (2014). Why time matters: codon evolution and the temporal dynamics of dN/dS. *Molecular Biology and Evolution*, 31(1), 212–231.
- Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., ... Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, 50(7), 920–927.
- Park, J. Y., An, Y.-R., Kanda, N., An, C.-M., An, H. S., Kang, J.-H., ... Kim, H. (2015). Cetaceans evolution: insights from the genome sequences of common minke whales. *BMC Genomics*, 16, 13.
- Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., ... Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344(6190), 1410–1414.
- Powell, R. W., & Kelly, W. (1977). Tool use in captive crows. *Bulletin of the Psychonomic Society*, 10(6), 481-483. doi: 10.3758/bf03337704
- Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology: CB*, 20(4), R208–R215.
- Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*, 13(6), 735-748. doi: 10.1093/oxfordjournals.molbev.a025634
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Retrieved from Available from <http://www.R-project.org/>
- Rutz, C., & Hunt, G. R. (2020). New Caledonian crows afford invaluable comparative insights into human cumulative technological culture. *The Behavioral and Brain Sciences*, 43, e177.

- Rutz, C., Klump, B. C., Komarczyk, L., Leighton, R., Kramer, J., Wischnewski, S., ... Masuda, B. M. (2016). Discovery of species-wide tool use in the Hawaiian crow. *Nature*, 537(7620), 403–407.
- Rutz, C., Ryder, T. B., & Fleischer, R. C. (2012). Restricted gene flow and fine-scale population structuring in tool using New Caledonian crows. *Die Naturwissenschaften*, 99(4), 313–320.
- Rutz, C., & St Clair, J. J. H. (2012). The evolutionary origins and ecological context of tool use in New Caledonian crows. *Behavioural Processes*, 89(2), 153–165.
- Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., ... Edwards, S. V. (2019). Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, 364(6435), 74–78.
- Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., ... Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and
- Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. JHU Press.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1), 23–35. doi: 10.1017/s0016672300014634
- Sprevak, M. (2014). *The Evolved Apprentice: How Evolution Made Humans Unique*. The Jean Nicod Lectures. By Kim Sterelny. A Bradford Book. Cambridge (Massachusetts): MIT Press. \$35.00. xvi 242 p.; ill.; index. ISBN: 978-0-262-01679-7. 2012. *The Quarterly Review of Biology*, Vol. 89, pp. 176–176. doi: 10.1086/676111
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- St Clair, J. J. H., Burns, Z. T., Bettaney, E. M., Morrissey, M. B., Otis, B., Ryder, T. B., ... Rutz, C. (2015). Experimental resource pulses influence social-network dynamics and the potential for information flow in tool-using crows. *Nature Communications*, 6, 7197.
- St Clair, J. J. H., Klump, B. C., Sugasawa, S., Higgott, C. G., Colegrave, N., & Rutz, C. (2018). Hook innovation boosts foraging efficiency in tool-using crows. *Nature Ecology & Evolution*, 2(3), 441–444.
- Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, 25(1), 79–88.
- Sterelny, K. (2012). *The evolved apprentice: how evolution made humans unique*. MIT press.
- Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2019). qvalue: Q-value estimation for false discovery

rate control. R package version 2.16.0. [Http://github.com/jdstorey/qvalue](http://github.com/jdstorey/qvalue).

- Suemoto, H., Muragaki, Y., Nishioka, K., Sato, M., Ooshima, A., Itoh, S., ... Fässler, R. (2007). Trps1 regulates proliferation and apoptosis of chondrocytes through Stat3 signaling. *Developmental Biology*, 312(2), 572–581.
- Tajima, F. (1993). Statistical analysis of DNA polymorphism. *Idengaku Zasshi*, 68(6), 567–595.
- Taylor, A. H., Elliffé, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural innovation in New Caledonian crows. *Proceedings. Biological Sciences / The Royal Society*, 277(1694), 2637–2643.
- Taylor, A.H., Gray, R.D. (2014) Is there a link between the crafting of tools and the evolution of cognition? *Wiley Interdisciplinary Reviews: Cognitive Science*, 5 (6), 693-703.
- Teschke, I., Wascher, C. A. F., Scriba, M. F., von Bayern, A. M. P., Huml, V., Siemers, B., & Tebbich, S. (2013). Did tool-use evolve with enhanced physical cognitive abilities? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1630), 20120418.
- Troscianko, J., von Bayern, A. M. P., Chappell, J., Rutz, C., & Martin, G. R. (2012). Extreme binocular vision and a straight bill facilitate tool use in New Caledonian crows. *Nature Communications*, 3(1), 1-7. doi: 10.1038/ncomms2111
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics / Editorial Board, Andreas D. Baxeavanis ... [et Al.]*, 43, 11.10.1–33.
- Weissensteiner, M. H., Pang, A. W. C., Bunikis, I., Höjjer, I., Vinnere-Petterson, O., Suh, A., & Wolf, J. B. W. (2017). Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Research*, 27(5), 697–708.
- Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., & Nielsen, R. (2007). Localizing recent adaptive evolution in the human genome. *PLoS Genetics*, 3(6), e90.
- Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews. Genetics*, 13(1), 59–69.
- Wolf, J. B. W., Künstner, A., Nam, K., Jakobsson, M., & Ellegren, H. (2009). Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection. *Genome Biology and Evolution*, 1, 308–319.
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews. Genetics*, 8(3), 206–216.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*,

24(8), 1586–1591.

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12), 496–503.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., ... Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215), 1311–1320.

### Data accessibility

*C. moneduloides* de novo assembly: NCBI accession number: VRTO000000000; BioProject ID: PRJNA557728

*C. moneduloides* annotation: datadryad (doi: <https://doi.org/10.5061/dryad.w0vt4b8m9>)

*C. woodfordi* RNAseq data: BioProject ID: PRJNA577979 - SRA submission SUB6432980

Umbrella project ID: PRJEB33755

Short read data: PRJEB33706, PRJEB33707

### Authors contributions

G.R.H., R.D.G, M.G.R., C.R. J.B.W., and N.J.G. conceived the project, and M.G.R., C.R. J.B.W., and N.J.G. coordinated and ran the work programme. G.R.H., H.A., R.F., C.R., and N.J.G contributed samples and/or assisted in sample preparation and permitting. N.D., V.E.K., R.A.W.W., D.P. and K.R. performed the bioinformatic analyses, under supervision from M.G.R, J.B.W., and N.J.G. N.D., V.E.K., R.A.W.W., G.R.H., C.R. and N.J.G contributed to writing the main body of the manuscript with input from all other authors. N.D., V.E.K., R.A.W.W., wrote the supplementary documentation and R.D.G., R.F., M.G.R., C.R., J.B.W.W., and N.J.G contributed to the interpretation and presentation of results in the main manuscript and supplementary documents.

### Figure and Table legends

**Table 1.** Candidate genes under putative positive selection when comparing *C. moneduloides* and/or *C. hawaiiensis* to other crow species. Abbreviations: sp. - species.

Reference	dataset	Gene	Full gene name	$\omega$	Species with $\omega > 1$
-----------	---------	------	----------------	----------	---------------------------

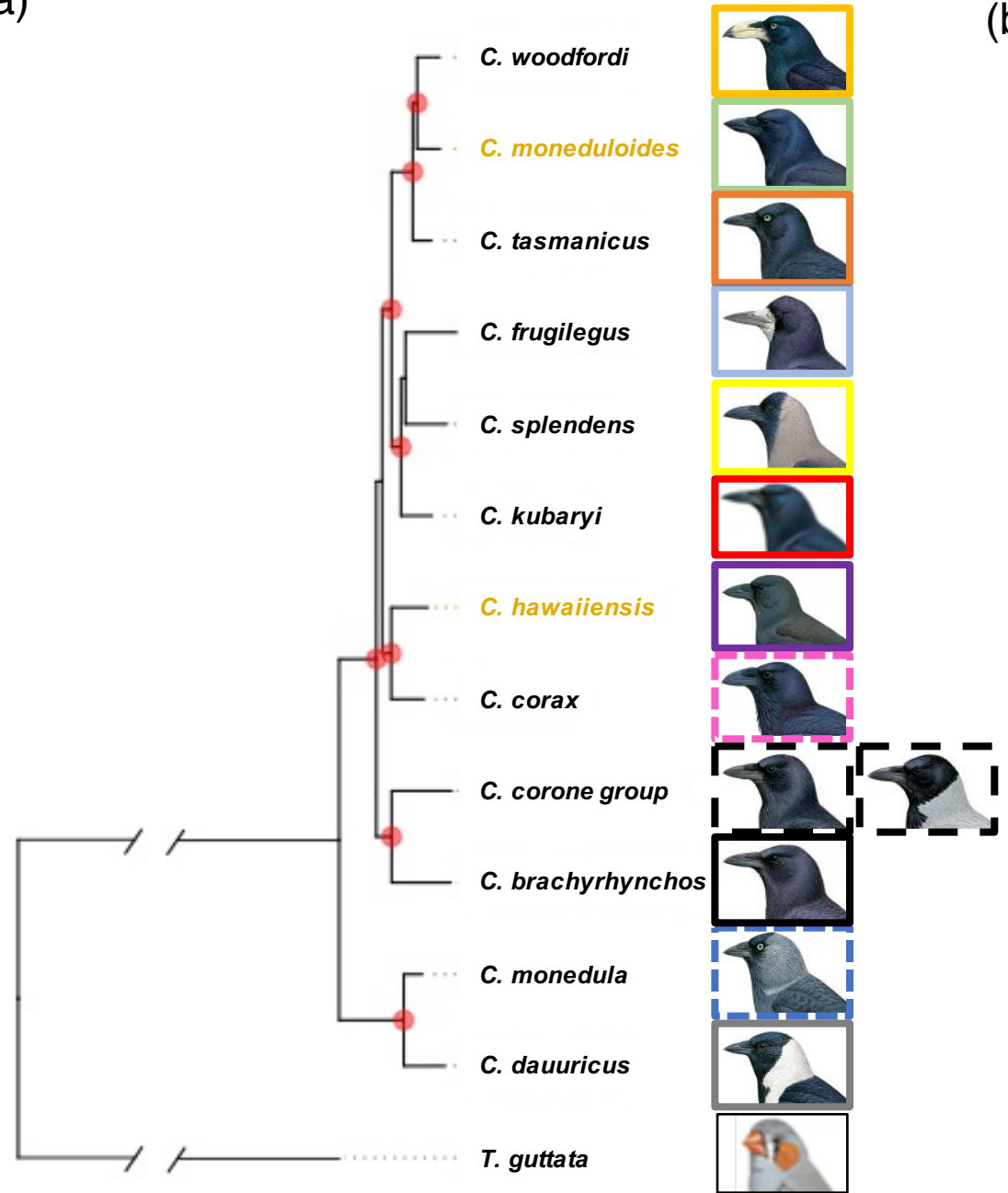
genome		name			
<i>C. moneduloides</i>	5 sp.	CAPN15	calpain-15 - Cysteine Peptidases	1.01	<i>C. moneduloides</i>
<i>C. moneduloides</i>	5 sp.	CAMLG	calcium signal-modulating cyclophilin ligand - Transporter	18.57	<i>C. moneduloides</i>
<i>C. moneduloides</i>	7 sp.	TPBGL	trophoblast glycoprotein-like	1.79	<i>C. moneduloides</i>
<i>C. moneduloides</i>	8 sp. same rates	CIITA	class II major histocompatibility complex transactivator	1.05	<i>C. moneduloides</i> and <i>C. hawaiiensis</i>
<i>C. moneduloides</i>	8 sp. diff. rates	SPSB1	splA/ryanodine receptor domain and SOCS box containing 1	1.46	<i>C. moneduloides</i>
<i>C. cornix</i>	5 sp.	LHX1	LIM homeobox 1	13.27	<i>C. moneduloides</i>
<i>C. cornix</i>	5 sp.	ZC2HC1A	zinc finger C2HC-type containing 1A	5.41	<i>C. moneduloides</i>
<i>C. cornix</i>	5 sp.	DTNBP1	dystrobrevin binding protein 1	1.78	<i>C. moneduloides</i>
<i>C. cornix</i>	7 sp.	HTRA2	HtrA serine peptidase 2	9.78	<i>C. moneduloides</i>
<i>C. cornix</i>	7 sp.	TBC1D10A	TBC1 domain family member 10A	7.99	<i>C. moneduloides</i>
<i>C. cornix</i>	8 sp. diff. rates	NPRL2	NPR2 like, GATOR1 complex subunit	2.48	<i>C. moneduloides</i>
<i>C. cornix</i>	8 sp. diff. rates	SLC6A20	solute carrier family 6 member 20	1.02	<i>C. moneduloides</i>

**Figure 1. Phylogeny and worldwide distributions of the *Corvus* species investigated. (a)** Maximum Likelihood Phylogeny based on 631 core eukaryotic, single-copy genes. Red points at nodes indicate  $\geq 95\%$  bootstrap support. Note that the branch to *T. guttata* has been truncated to 30% of its actual length for clearer visualisation. **(b)** Contemporary geographical distribution of crows used in this study.

**Figure 2. Genome-wide diversity in *C. moneduloides*. (a)** Nucleotide diversity ( $\square$ ) and tests of neutrality summary statistics (Fay and Wu's  $H$ , and Tajima's  $D$ ) in windows across all *C. (c.) cornix* scaffolds, placed on *in silico* chromosomes from Weissensteiner et al. (2017). **(b)** Tajima's  $D$  statistics in the 11 regions identified as candidates for positive selection as well as the locations of five genes with a known association to development of bill morphology in birds (depicted with black horizontal bars). Regions highlighted by vertical

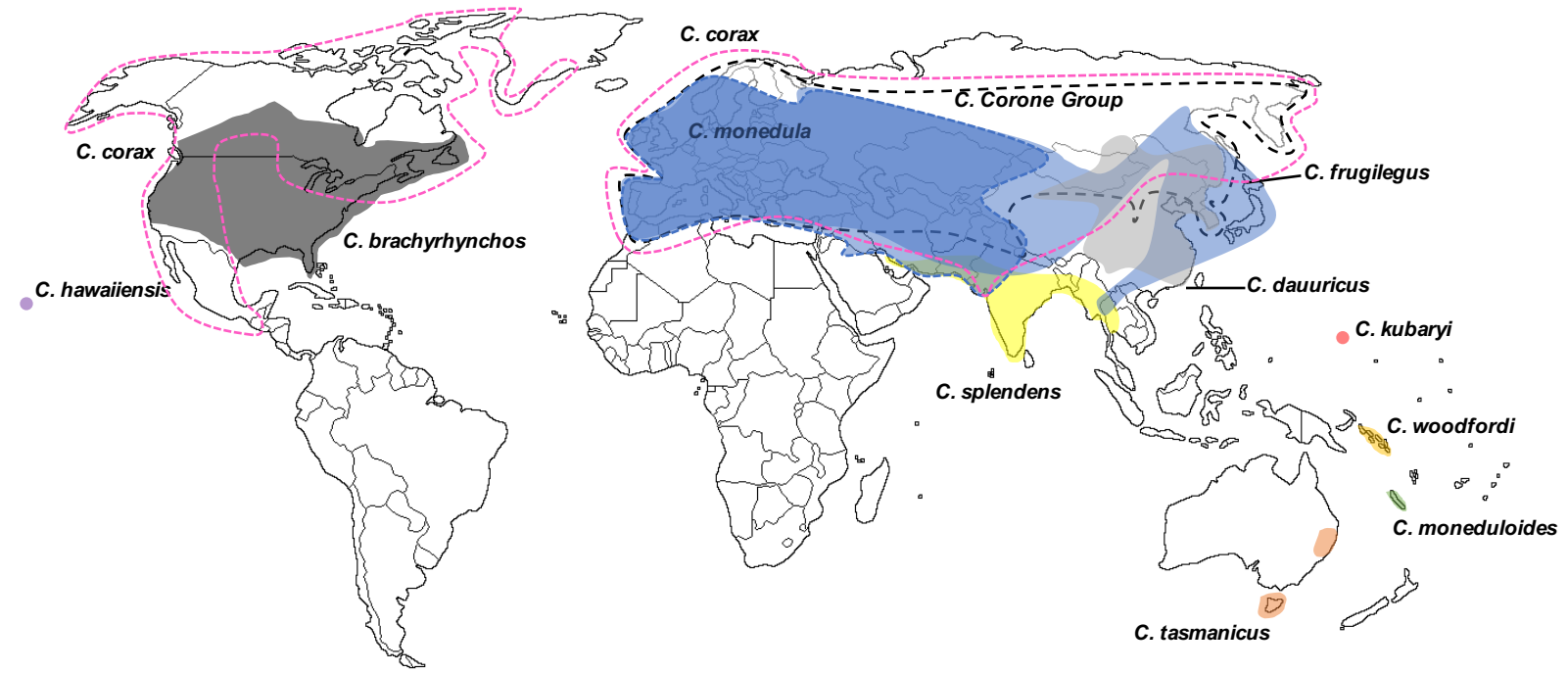
grey regions show where Tajima's  $D$  is below the theoretical threshold of -1.51 (dashed horizontal lines) in *C. moneduloides* but not in *C. woodfordi*.

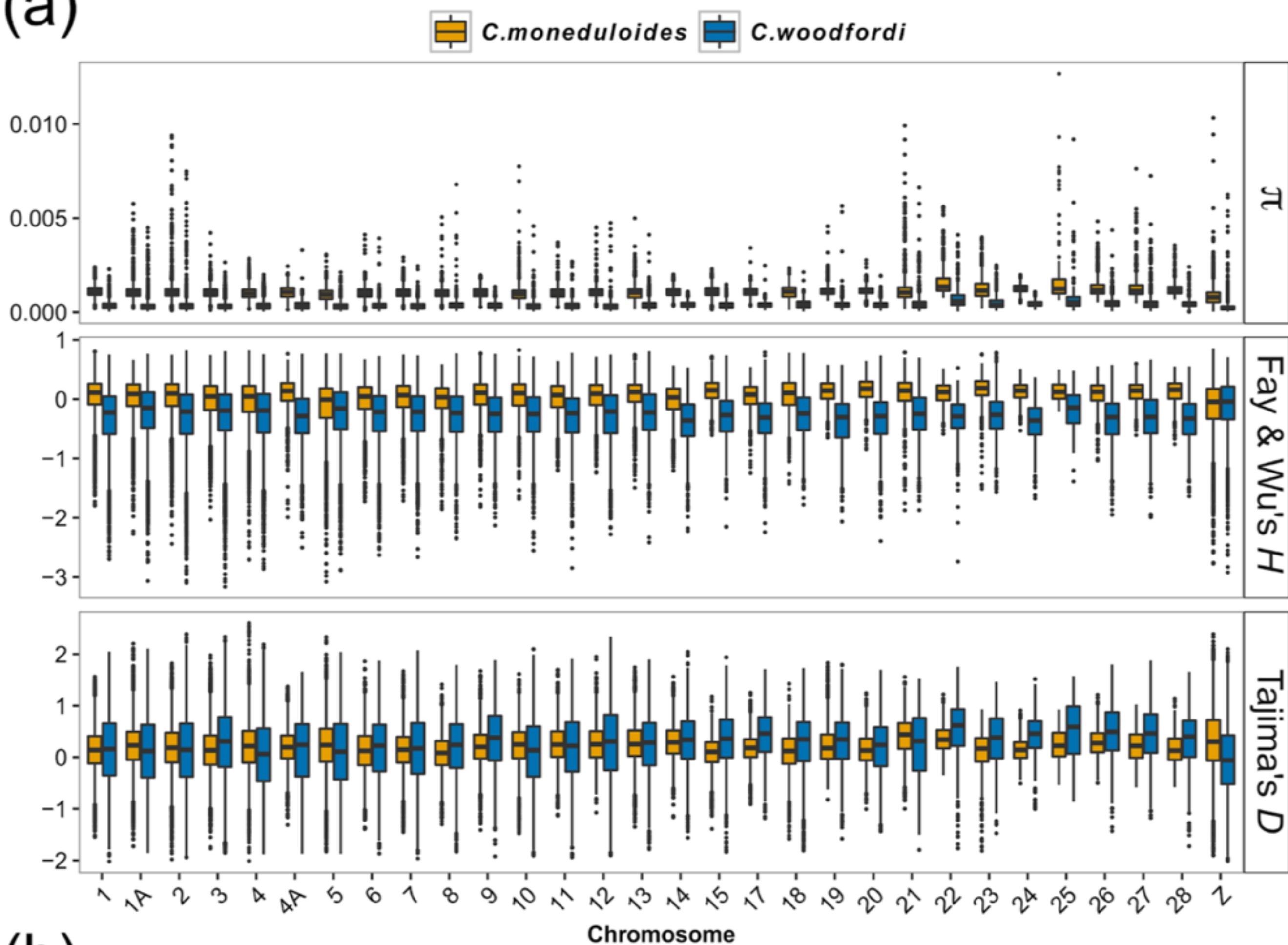
(a)



0.01

(b)



**(a)****(b)**