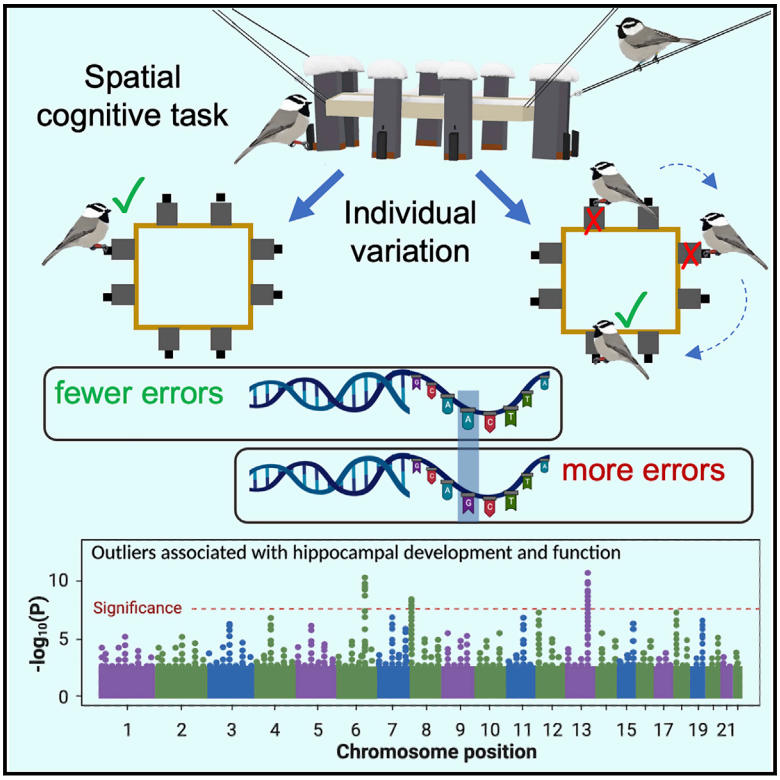
Report

Current Biology

The genetic basis of spatial cognitive variation in a food-caching bird

Graphical abstract



Highlights

- Food-caching birds use spatial cognition to recover food stores and survive winter
- Variation in cognitive phenotypes is associated with variation across the genome
- Top outlier genes are associated with hippocampal development and function
- Results link cognitive and genetic variation, making it available for selection

Authors

Carrie L. Branch, Georgy A. Semenov, Dominique N. Wagner, ..., Eli S. Bridge, Scott A. Taylor, Vladimir V. Pravosudov

Correspondence

clb345@cornell.edu (C.L.B.), vpravosu@unr.edu (V.V.P.)

In brief

Branch et al. find that individual variation in spatial cognition is associated with variation across the genome, showing that top outlier genes are associated with hippocampal development and function. This work is conducted on nonmigratory birds that use spatial learning and memory to relocate their food stores and survive harsh winters.

Branch et al., 2022, Current Biology *32*, 1–10
January 10, 2022 Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2021.10.036

CellPress

Current Biology

CellPress

Report

The genetic basis of spatial cognitive variation in a food-caching bird

Carrie L. Branch,[1,5,*] Georgy A. Semenov,[2,5] Dominique N. Wagner,[2] Benjamin R. Sonnenberg,[3] Angela M. Pitera,[3] Eli S. Bridge,[4] Scott A. Taylor,[2,6] and Vladimir V. Pravosudov[3,6,7,*]

[1]Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA
[3]Ecology, Evolution, and Conservation Biology Graduate Program, University of Nevada, Reno, NV 89557, USA
[4]Ecology and Evolutionary Biology, University of Oklahoma, Norman, OK 73019, USA
[5]These authors contributed equally
[6]These authors contributed equally
[7]Lead contact
*Correspondence: clb345@cornell.edu (C.L.B.), vpravosu@unr.edu (V.V.P.)
https://doi.org/10.1016/j.cub.2021.10.036

SUMMARY

Spatial cognition is used by most organisms to navigate their environment. Some species rely particularly heavily on specialized spatial cognition to survive, suggesting that a heritable component of cognition may be under natural selection. This idea remains largely untested outside of humans, perhaps because cognition in general is known to be strongly affected by learning and experience.[1–4] We investigated the genetic basis of individual variation in spatial cognition used by non-migratory food-caching birds to recover food stores and survive harsh montane winters. Comparing the genomes of wild, free-living birds ranging from best to worst in their performance on a spatial cognitive task revealed significant associations with genes involved in neuron growth and development and hippocampal function. These results identify candidate genes associated with differences in spatial cognition and provide a critical link connecting individual variation in spatial cognition with natural selection.

RESULTS AND DISCUSSION

Inter- and intra-specific variation in cognitive abilities and associated brain morphology is a topic of great interest and debate because the exact mechanisms of their evolution remain elusive.[1–3] Cognition in particular is thought to be primarily affected by environmental events including development and individual experience,[4] yet the contribution of genetic variation to naturally occurring cognitive variation remains unclear. It has long been hypothesized that the remarkably specialized spatial cognitive abilities required for caching and recovering food stores have evolved across species via natural selection (e.g., Figure 1A).[1,5,6] However, until recently, the evidence supporting this hypothesis was indirect. For example, birds experiencing harsher winter conditions, where reliance on food caches is greater (e.g., chickadee species at higher elevations and latitudes; Figure 1B), have been shown to outperform those inhabiting milder environments on spatial cognitive tasks[7,8,9] and exhibit significant differences in hippocampus morphology, including increased neuron number, soma size (Figure 1C), and hippocampal neurogenesis rates.[7,8] Moreover, differences in spatial cognition have been associated with differential gene expression in the hippocampus,[10] and hybrid chickadees (F1 and backcrossed hybrids) show deficiencies in both learning and memory, conceivably due to a breakdown of relevant genetic pathways.[11,12] Together, this work warrants investigation into possible genetic mechanisms underlying variation in spatial cognition in natural populations.

Our recent research on wild, free-living mountain chickadees (*Poecile gambeli*) inhabiting an elevation gradient in the northern Sierra Nevada provided the first direct evidence for natural selection via differential survival based on spatial learning and memory abilities.[14] Mountain chickadees are non-migratory food-caching birds that rely on specialized spatial memory to recover thousands of stored food items scattered throughout their territories.[6] Individuals with better spatial learning and memory abilities are more likely to survive their first winter compared to those with worse spatial cognition.[14] Additionally, spatial cognitive ability does not appear to change with age and experience, suggesting that it is a temporally stable trait.[14–16] Given the evidence that natural selection acts on individual variation in spatial cognitive abilities of food-storing birds, this variation should be heritable and hence should have a genetic basis.[17,18]

Although spatial cognitive abilities are critical for overwinter survival in food-caching birds, considerable variation persists both within and among populations. In mountain chickadees in the northern Sierra Nevada, this is likely due to gene flow along an elevation gradient,[19] temporally variable selection on a complex trait, and, perhaps, certain environmental influences on cognition (Figures 1A and 1B). Our long-term spatial cognitive dataset collected using well-established radio frequency identification (RFID) enabled "smart" feeders (Figure 1D; Video S1)

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

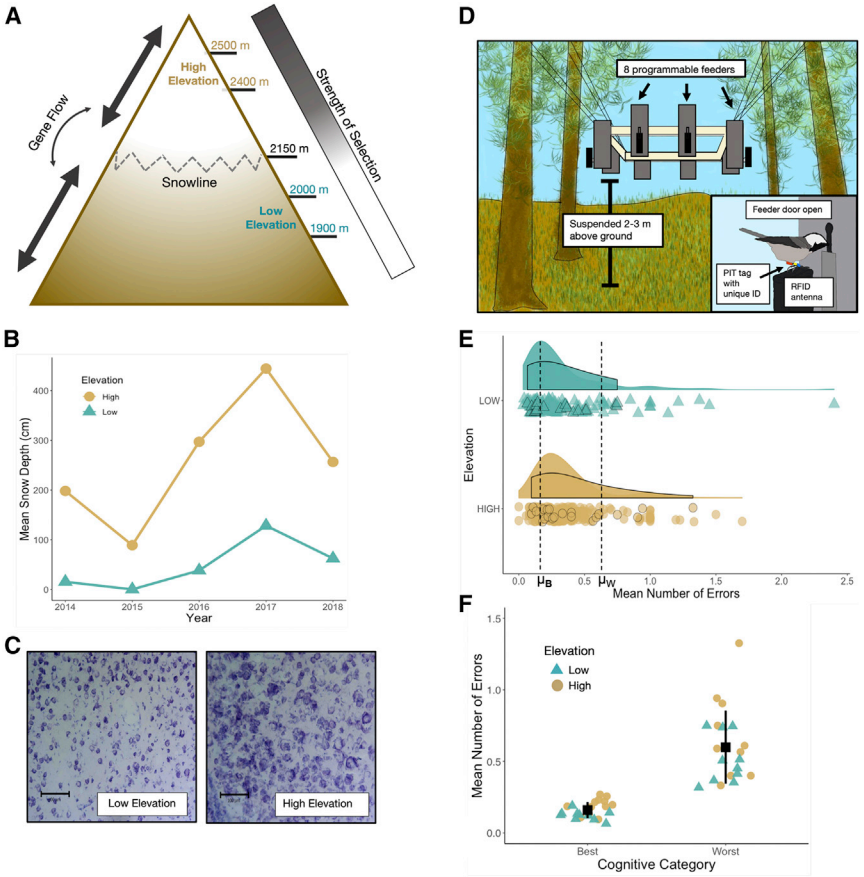


Figure 1. Testing spatial cognition in the wild

(A) Schematic depicting high and low elevation sites used for sampling, including effects of gene flow (thickness of arrows represent high rates of effective migration above and below the snowline, with less movement across the snowline) and strength of selection (white to black gradient represents increased strength of selection).

(B) Example of annual variation in mean snow depth (cm) for high and low elevation sites across a total of 5 years, including 3 years used in the study (2016–2018). Mean snow depth is from April 1st of each year and collected from SNOTEL weather stations at Sagehen Experimental Forest, CA, USA.

(C) Previously documented significant individual differences in hippocampal neuron number and soma size in our study population.[7]

(D) Schematic of "smart" spatial cognition testing apparatus with inset showing an individual bird on the RFID reader retrieving a seed with the mechanical door open.

(E) Distribution of spatial cognitive performance for high (n = 243) and low (n = 162) elevation birds across 3 years of testing (2016–2018). Individual performance scores and total distribution of individuals selected for genomic analyses are outlined in black. Note that the distribution of selected birds spans the entire range of performance. Mean of best (μ_B) and worst (μ_W) groups used in our analyses represented by dotted lines.

(F) Spatial cognitive performance of birds selected for GWAS from (E). Black squares and whiskers represent mean and standard deviation. Individual points are adjusted to avoid overlap.[13] High elevation birds are represented by tan circles and low elevation birds are represented by teal triangles within best and worst cognitive category. Related to Data S1.

provides an unprecedented opportunity to investigate the genetic basis of this naturally occurring variation in spatial cognitive abilities.

We used whole-genome sequencing, combining traditional genome-wide association studies (GWASs) and a Random Forest machine learning approach,[20] to compare the genomes of wild, free-living birds. We sampled birds from high and low elevations that performed the best on a spatial cognitive task, all of whom survived more than one year (n = 22), to those that performed the worst on the task and generally did not survive more than 1 year (n = 15/20)—the group with better spatial cognition was associated with a significant survival advantage (Fisher's exact test, $p < 0.001$). Birds from both high and low elevations were selected for each performance group to ensure that the strongest signal between groups was variation in cognition and not a correlate of elevation.

Spatial cognitive ability was assessed using automated "smart" feeder arrays equipped with RFID technology.[14,15,21,22] Birds were fitted with plastic leg bands embedded with passive integrative transponder (PIT) tags, allowing for unique identification. The RFID spatial array consisted of 8 equidistant feeders programmed so that each bird was only rewarded at one feeder location, while all feeders recorded the identity and time of visit by any PIT-tagged birds (Figure 1D; Video S1). Spatial cognitive ability was measured as the mean number of errors (number of unrewarding feeders visited prior to visiting the correct rewarding feeder) per trial across the entire testing period. This measure provides a robust representation of ecologically relevant spatial cognitive abilities because it is repeatable within individuals[14,15] and is associated with fitness consequences, both direct via differential survival[14] and indirect via differential female reproductive investment.[22]

We sampled 42 mountain chickadees across 3 years of testing from the extremes of the cognitive performance range: 22 were chosen as the best and 20 were chosen as the worst. Performance scores of individuals in the best and worst groups did not overlap, but individual variation within each group provided a continuous distribution from best to worst (Figure 1F). We intentionally chose individuals with means in the tails of the cognitive performance distribution (Figures 1E and 1F) to amplify the signal of genetic associations, although we acknowledge that this design could inflate associations from loci with the largest effect to the detriment of small-effect polygenes. There was a significant difference in the mean number of errors per trial between best (mean errors/trial: 0.16 ± 0.045) and worst (0.60 ± 0.05) performers (cognitive category [best versus worst]: $F_{1,38} = 72.91$, $p < 0.0001$; Figure 1F), but there was not a significant effect of elevation, as we selectively picked the best and worst

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

CellPress

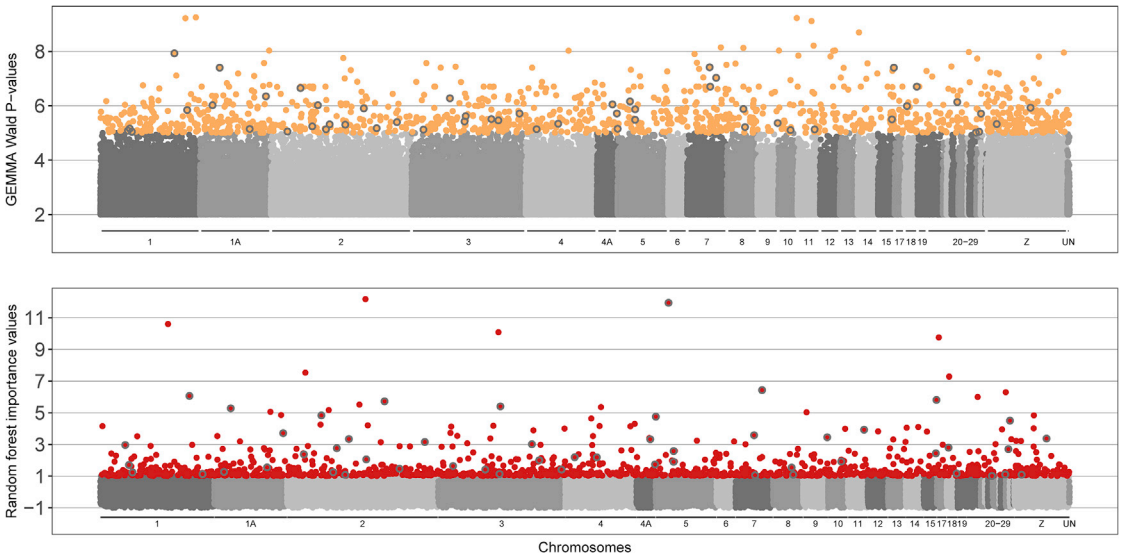


Figure 2. Genetic associations with the spatial cognition phenotype

Top: genome-wide association study results (GEMMA) with outliers exceeding −log10(p) > 5 highlighted in orange. Bottom: importance values for individual SNPs inferred using a machine learning algorithm (Random Forest), with the top 0.1% of loci highlighted in red. Dark outlines on both plots are outliers matching between GEMMA and Random Forest.

Related to Table S2.

performers at each elevation (elevation [high versus low]: $F_{1,38}$ = 1.38, p = 0.247; total trials completed [covariate]: $F_{1,38}$ = 5.26, p = 0.028).

We sequenced whole genomes of these individuals and identified ~41 million SNPs after alignment to the closely related black-capped chickadee (*P. atricapillus*) reference genome (https://www.ncbi.nlm.nih.gov/bioproject/589043; accession: JAAMOC000000000[12]), as there is no mountain chickadee reference genome available. This SNP dataset was further filtered to 12,106,779 SNPs for our GWAS analysis and 1,312,917 SNPs for our Random Forest analysis. For the response variable in genomic analyses, we used the continuous distribution of individual performance (mean number of errors per trial for each bird, as described above) across both groups, rather than a discrete between-group approach.

Using genome-wide efficient mixed model analysis (GEMMA),[23] we identified 1,338 (p value of the Wald likelihood ratio test −log10(p) > 5) and 305 (−log10(p) > 6) SNPs associated with the cognitive phenotype (with 5 outliers of −log10(p) > 9) (Figure 2). Given that our genomic dataset consisted of ~12 million loci, the expected numbers of false positives under the above significance thresholds are 120 and 12 loci, respectively, indicating that our analysis picks up a meaningful signal of the genotype-phenotype associations. Most strongly significant associations were represented by single physically unlinked SNPs (rather than clusters of loci), consistent with a polygenic genetic architecture of the phenotype. Next, we assessed the genetic architecture and heritability of the spatial cognition phenotype using Bayesian sparse linear mixed models (BSLMM).[24] BSLMM accounts for linkage between loci and relatedness among individuals using a Bayesian, Markov Chain Monte Carlo approach. The non-zero effect of each locus on the phenotype is estimated, and a posterior probability distribution for the total percent variation explained (PVE) across all loci is reported.[24,25] The spatial cognitive phenotype appears to be heritable with a posterior PVE estimate of 92% (SD = 14.9%) when performed on all loci. Despite the apparent polygenic nature of cognitive performance differences, our BSLMM results suggested the major phenotypic effects may be governed by ~10 loci (n_gamma parameter of GEMMA, SD = 29.6) which accounted for 87% (PGE parameter of GEMMA, SD = 19%) of phenotypic variance. To assess the predictive power of BSLMM, we performed leave-one-out cross validation. This analysis revealed little ability of predicting an individual's phenotype based on its genotype (r^2 = 0.0027, p = 0.74; Figure S1), likely due to a modest sample size unable to provide predictive power. Partial dominance and epistasis can create a non-linear relationship between genotype and phenotype even in traits with a simple genetic architecture (e.g., Semenov et al.[26]), and such effects are expected to accumulate rapidly with an increasing number of causal variants. It is important to note that BSLMM may overestimate the heritability of phenotypes;[25] combined with our rather small sample size, these results should be interpreted with caution. Nevertheless, even if our estimates are inflated, they do suggest at least moderate to high heritability of the spatial cognitive phenotype.

Using data from the same 42 individuals, we also performed a Random Forest regression analysis using individual cognitive phenotype as a continuous response variable. Machine learning approaches, like Random Forest, consider combinations of multiple loci that may influence a single phenotype,[20] providing valuable information complementary to GWAS, such as GEMMA. This analysis assigned positive variable importance scores to 116,059 SNPs (averaged across 3 replicates), of which the top ~0.1% (n = 1,312) had the strongest (and most consistent across replicates) signal, suggesting a likely contribution of these loci in explaining the cognitive phenotype. In further support of the

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

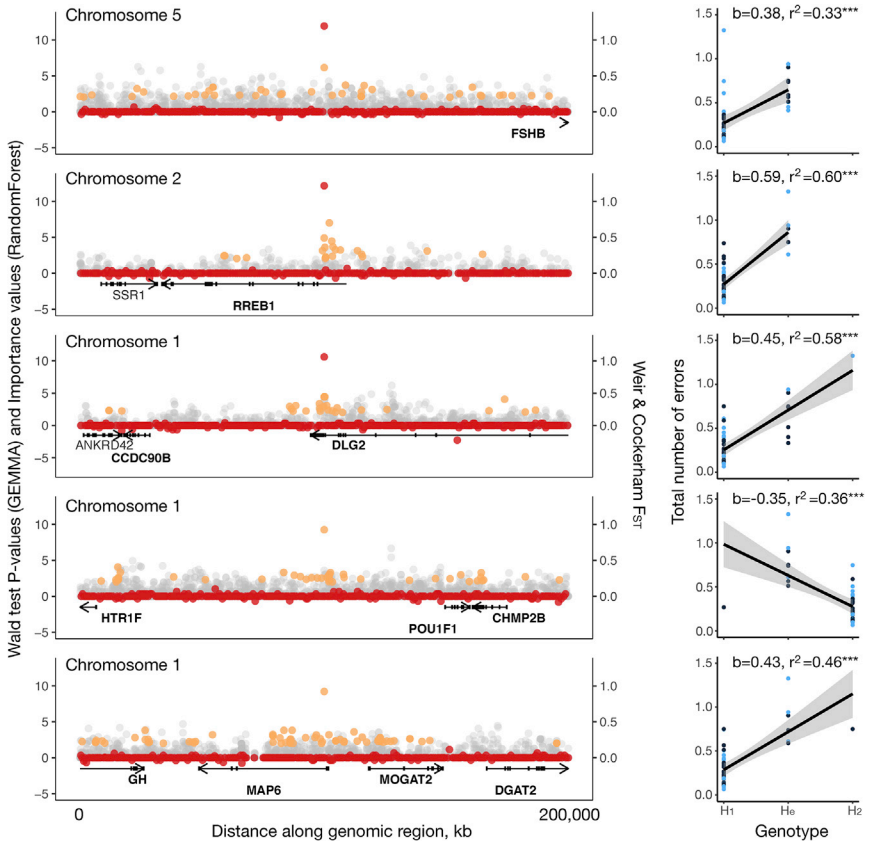


Figure 3. Genes in regions with highest association signal and genotype-phenotype relationship for significant SNPs

Left: zoom into a subset of genomic regions (200,000-bp total length of each), containing some of the most significant associations of GEMMA (orange), Random Forest (red), or both. Gray dots are single-locus FST. Black lines are gene models with boxes indicating exons, and arrows show the direction of reading frame. Indicated in bold are names of genes with known behavioral or neurological function or/and connected to behavioral disorders. Two genes with putative connection to behavior (ANKRD42 and SSR1, known to express in nervous system) are shown in regular font. Right: relationships between genotypes at SNPs with the highest significance in each region and cognitive phenotype (total number of errors). H1, H2, and He stand for homozygotes 1 and 2 and heterozygotes, respectively. b is the slope of linear regression and gray shading is its 95% CI. *** indicate significance under 0.0001. r^2 is adjusted R-squared. Dark and light blue colors indicate males and females, respectively. Note that there is an apparent lack of "worst" homozygotes, suggesting that carriers of such genotypes did not survive to adulthood and hence were not sampled.
Related to Figures S3 and S4.

likely role of these loci in determining cognitive differences, 52 loci were identified as highly significant outliers by both GEMMA and Random Forest (Figure 2), despite the drastically different algorithms implemented in these analyses.

Using the above outlier loci, we identified 1,251 and 1,225 known genes for GEMMA and Random Forest analyses, respectively, of which 766 and 937 were unique (Table S2). Strikingly, 266 genes overlapped between GEMMA (34% of total genes) and Random Forest (28% of total genes) analyses. Using the PANTHER classification system (http://pantherdb.org/), we examined the list of genes from both the regression GWAS (GEMMA) and Random Forest analyses for statistical overrepresentation (PANTHER Gene Ontology [GO]-Biological Process Complete) with a Bonferroni correction and the human and chicken genome annotations (Tables S1 and S3). While the chicken genome is taxonomically closer to the chickadee genome, the human genome has far more detailed information about gene models and functions. PANTHER revealed many GO categories with significant positive overrepresentation and connection to neurological function and development of the nervous system (Table S1), including neuron growth and development, telencephalon development, and neurogenesis (Figure S2). Interestingly, the top GO category for Random Forest using the chicken genome as a reference was "regulation of alkaline phosphatase activity" (Table S3), and alkaline phosphatase is known to promote axonal growth of hippocampal neurons.[27] Overall, the Developmental Process GO category was associated with 41% of all outlier genes identified by both GEMMA and Random Forest using the human genome as a reference (35.5% GEMMA and 39.5% Random Forest using chicken genome). The Nervous System Development GO category was associated with 20.9% (GEMMA) and 19.4% (Random Forest) of all outlier genes using human genome (16.2% GEMMA and 19.7% Random Forest using chicken genome).

We further looked at the distribution of outliers for a subset of genes with the highest association values of GEMMA, Random Forest, or both (Figures 3 and S3). Individual outlier loci showed variable patterns of association with gene features and were located within protein-coding regions and in non-coding regions in the vicinity of genes (Figures 3 and S3). All of these genes are particularly strong candidates for differences in cognitive performance due to known functions in hippocampus development and function including neurogenesis and associations with behavioral disorders (see Table 1 for gene functions and references). Several genes in our association datasets have well known associations with the development of the nervous system, the brain, and the hippocampus. These are of particular interest given that the hippocampus (and its taxon-specific homologs) is the area of the brain associated with spatial learning and memory.[6] These include ROBO1, ROBO2, and SLIT2—genes known for their involvement in axon guidance, brain development, progenitor cell proliferation, and migration[28]—as well as WNT3A, LEF1, and ZEB2, which are critical for development of the hippocampus.[29,30] Additionally, FGF13 is essential for hippocampal neurogenesis in rodents;[31] CNTN6 is critical for brain and hippocampus development, and it has been implicated as an autism risk gene;[32] BMP2 is known to affect hippocampal function associated with learning and memory;[33] and AGAP3 is involved in regulating synaptic strength associated with learning

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

Table 1. Top ranked genes and known cognitive relevance

Gene name	Description	References
HTR1F	involved in learning and memory, also reported in chickens	37
POU1F1	linked to cognitive delays	38
CHMP2B	associated with cognitive impairments	39
GH	involved in cognitive function both in mammals and in fish	40,41
MAP6	involved in cognitive function, neuron development and maturation, synaptic plasticity, and brain connectivity	42,43
MOGAT2	linked to brain and hippocampal function associated with Alzheimer's disease	44
DGAT2	linked to hippocampal function associated with glucocorticoid responsiveness	45
FSHB	involved in brain development	46
RREB1	linked to hippocampus-related psychiatric disorders	47
CCDC90B	linked to hippocampal function	48
DLG2	linked to neural development and cognitive function	49
SSR1	linked to apoptosis in zebrafish	50
ANKRD42	linked to brain-related complex traits including intelligence	51
SESTD1	involved in development of the brain and the hippocampus	52
IGSF11	involved in synaptic plasticity in the hippocampus	53
RTN4R	linked to cognitive function	54
TNC	involved in neurogenesis	55
NEO1	involved in hippocampal function	56
CCNE1	involved in cell proliferation in humans (and also detected in birds)	57
VSTM2B	linked to dementia	58
POP4	linked to hippocampal function, aging, and Alzheimer's disease	59
C19 or F12	linked to cognitive decline and neuropsychiatric impairment	60
METRN	involved in neuron development and neurogenesis	61
WDR24	involved in cell growth and metabolism	62
RHBDL1	involved in development	63
ZNF385B	linked to neurodevelopmental disorders	64
ARHGAP31	linked to brain development	65
ZFPM2	involved in neuron development in the brain and the hippocampus	66
NME3	involved in hippocampal function	67
SPSB3	involved in hippocampal function	68
IGFALS	involved in hippocampal function	69
JMJD8	linked to cell metabolism, cell proliferation, and apoptosis	70
PLEK	linked to hippocampus development and function	71
CFAP70	linked to brain function related to Alzheimer's disease	44
MSS51	linked to neurogenesis and intellectual ability	72
RHOT2	linked to hippocampal plasticity	73
ANTKMT	also known as FAM173A, involved in hippocampal f unction and memory	74
EIPR1	involved in neural functions associated with psychiatric disorders	75
UPK1B	involved in hippocampal function	76

Description and references of the top ranked genes based on the highest association values of GEMMA, Random Forest, or both. Related to Figure S3 and Table S2.

and memory.[34] Some of the genes we detected (DRD2, NMDE2 [NR2B]) have also been specifically implicated in cognitive function in birds.[35,36]

We also detected two genes (GRM3 and ELMO1) that exhibit differential expression in the hippocampus between black-capped chickadee populations at different latitudes.[10] These two black-capped chickadee populations also show differences in hippocampal morphology and spatial learning and memory (better spatial cognition, larger number of hippocampal neurons, and higher rate of adult hippocampal neurogenesis in Alaska chickadees compared to Kansas chickadees),[10] like the mountain chickadees studied here. Differential expression of these genes was detected in birds that were hand-reared from when they were 10 days old and maintained in identical lab conditions, which suggests these differences are associated with genetic variation rather than experience. That we detected significant associations between these two genes and cognitive performance in the present study suggests that these differences are functional. GRM3 is well known to be associated with hippocampal dependent function, including memory in humans,[77] and ELMO1 is involved in neuron development and adult neurogenesis.[78,79]

Finally, there were significant associations between genotypes and the cognitive phenotype at a series of individual SNPs (best associations of GEMMA, Random Forest, or both) we examined (Figures 3 and S4). The strength of association was low to moderate (r^2 = 0.17-0.61), which follows expectations for a highly polygenic trait. Interestingly, in all but one case examined (n = 14), there was an apparent lack of homozygous genotypes for the "worst" allele (Figures 3 and S4). This result suggests that carriers of these homozygous genotypes might experience behavioral or physiological disadvantages and rarely survive to adulthood (and hence were not sampled in our dataset) and warrant further study. Overall, the fact that every SNP

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

tested for our focal gene set exhibited a significant association with cognitive phenotype suggests that these SNPs are important for spatial cognition and that they contribute to individual variation in cognitive phenotype. In half of the top SNPs analyzed (7 of 14; Figures 3, S3, and S4) the strength of association with cognitive phenotype was above 50%, suggesting at least moderate heritability, consistent with our BSLMM results.

Overall, using our unique long-term field dataset, we found significant differences in numerous genes associated with individual variation in spatial learning and memory ability in food-caching mountain chickadees. Many of these differences are associated with development of the nervous system and the brain, including neuron growth and development, and neurogenesis. While adult neurogenesis is known to occur in avian song nuclei and in the olfactory system, hippocampal neurogenesis consisting of proliferation, survival, and incorporation of new neurons has received the most attention because of its direct connection to general cognitive function and spatial learning and memory in mammals and birds.[80–83] The genes that differ among the individuals varying in their spatial cognitive abilities in our study population are also associated with nervous system development and function, suggesting a genetic basis or constraint on the natural variation present in the spatial memory and learning abilities of food-caching species. In other words, an individual's genes appear to lay the cognitive foundation upon which that individual may then build via learning and experience.[84] This suggests that the chickadee spatial cognition phenotype may be determined, at least in part, early in development,[84] but more study is required to firmly establish if certain genes or loci are predictive of spatial cognition in chickadees. Our results are consistent with previous studies showing that differences in the spatial cognition of chickadees are associated with differences in the number and size of hippocampal neurons (Figure 1C), as well as adult hippocampal neurogenesis rates.[7,8] In addition to the genes involved in development, many of the genes identified appear to be directly involved in neuronal and hippocampal functions associated with learning and memory. Combined with previous results showing differential survival of chickadees based on their spatial cognitive abilities,[14] our finding that differences in spatial cognition have a genetic basis indicates that natural selection on spatial cognition can result in local adaptation.

Our results provide an important contribution to the understanding of the genetic basis of naturally occurring behavioral and cognitive variation. Often behavior, and more specifically, cognition, is thought of as labile and readily affected by experience.[4] Historically, much of the work addressing variation in cognition was limited to lab environments and humans, which comes with a host of biases; however, advances in technology and genomic analytics have paved the way to test many long-standing assumptions in ecology and evolutionary biology, resulting in a better understanding of naturally occurring phenotypic variations. Our study provides the first direct evidence for the critical component expected for the evolution of cognitive variation by natural selection, clearly demonstrating that genetic differences underlie natural individual variation of spatial cognitive abilities in a wild population of birds.

The results of our study lay the foundation for future investigations into the genetic basis of, and natural selection on, spatial cognitive abilities in wild food-caching birds. Future work will focus on quantifying heritability using a larger and more representative sample, investigating the impact of strong selective events on allele frequencies, and understanding temporal variation and geographic consistency in selection on spatial learning and memory.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study subjects and site
- METHOD DETAILS
  - Cognitive testing in the wild
  - Library preparation and whole genome sequencing
  - Whole genome alignment and filtering
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Measuring cognitive ability
  - Regression GWAS analysis
  - Random Forest filtering and analysis
  - Gene ontology analysis
  - Gene annotation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.10.036.

ACKNOWLEDGMENTS

Thank you to Drs. Rebecca Croston and Maria Tello-Ramos for their contribution to data collection and the staff of the Sagehen Creek Field Station (University of California Berkeley) for assistance at our field site. We acknowledge that this research was conducted on the ancestral and traditional lands of the Washoe people. We recognize and support the Indigenous individuals who live there now and those forcibly removed from their homelands (https://washoetribe.us). This research was supported by NSF grants IOS 1856181 and IOS 1351295 to V.V.P. C.L.B. is supported by a Rose Postdoctoral Fellowship from the Cornell Lab of Ornithology. G.A.S. is supported by NSF grant RoL 1928891. We also thank 3 anonymous reviewers for their constructive criticism and suggestions that have greatly improved the manuscript.

AUTHOR CONTRIBUTIONS

C.L.B., V.V.P., and S.A.T. conceived and designed the experiment. C.L.B., B.R.S., A.M.P., and V.V.P. collected cognitive field data. C.L.B. and V.V.P. analyzed the cognitive data. E.S.B. and V.V.P. co-designed the "smart" feeder arrays. E.S.B. designed RFID boards and wrote the software. G.A.S., D.N.W., and S.A.T. conducted all genomic methods and analyses: GWAS, Random Forest, and GO analyses. C.L.B., G.A.S., D.N.W., S.A.T., and V.V.P. co-wrote the manuscript. All authors provided feedback for the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: July 8, 2021
Revised: September 15, 2021
Accepted: October 14, 2021
Published: November 3, 2021

REFERENCES

1. Sherry, D.F., Vaccarino, A.L., Buckenham, K., and Herz, R.S. (1989). The hippocampal complex of food-storing birds. Brain Behav. Evol. *34*, 308–317.
2. Morand-Ferron, J., and Quinn, J.L. (2015). The evolution of cognition in natural populations. Trends Cogn. Sci. *19*, 235–237.
3. Morand-Ferron, J. (2017). Why learn? The adaptive value of associative learning in wild populations. Curr. Opin. Behav. Sci. *16*, 73–79.
4. Shettleworth, S.J. (2010). Cognition, Evolution and Behavior, Second Edition (Oxford University Press).
5. Krebs, J.R., Sherry, D.F., Healy, S.D., Perry, V.H., and Vaccarino, A.L. (1989). Hippocampal specialization of food-storing birds. Proc. Natl. Acad. Sci. USA *86*, 1388–1392.
6. Pravosudov, V.V., and Roth, T.C., II. (2013). Cognitive ecology of food hoarding: The evolution of spatial memory and the hippocampus. Annu. Rev. Ecol. Evol. Syst. *44*, 173–193.
7. Freas, C.A., LaDage, L.D., Roth, T.C., II, and Pravosudov, V.V. (2012). Elevation-related differences in memory and the hippocampus in mountain chickadees, *Poecile gambeli*. Anim. Behav. *84*, 121–127.
8. Roth, T.C., 2nd, LaDage, L.D., Freas, C.A., and Pravosudov, V.V. (2012). Variation in memory and the hippocampus across populations from different climates: a common garden approach. Proc. Biol. Sci. *279*, 402–410.
9. Croston, R., Kozlovsky, D.Y., Branch, C.L., Parchman, T.L., Bridge, E.S., and Pravosudov, V.V. (2016). Individual variation in spatial memory performance in wild mountain chickadees from different elevations. Anim. Behav. *111*, 225–234.
10. Pravosudov, V.V., Roth, T.C., 2nd, Forister, M.L., Ladage, L.D., Kramer, R., Schilkey, F., and van der Linden, A.M. (2013). Differential hippocampal gene expression is associated with climate-related natural variation in memory and the hippocampus in food-caching chickadees. Mol. Ecol. *22*, 397–408.
11. McQuillan, M.A., Roth, T.C., 2nd, Huynh, A.V., and Rice, A.M. (2018). Hybrid chickadees are deficient in learning and memory. Evolution *72*, 1155–1164.
12. Wagner, D.N., Curry, R.L., Chen, N., Lovette, I.J., and Taylor, S.A. (2020). Genomic regions underlying metabolic and neuronal signaling pathways are temporally consistent in a moving avian hybrid zone. Evolution *74*, 1498–1513.
13. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
14. Sonnenberg, B.R., Branch, C.L., Pitera, A.M., Bridge, E., and Pravosudov, V.V. (2019). Natural selection and spatial cognition in wild food-caching mountain chickadees. Curr. Biol. *29*, 670–676.e3.
15. Tello-Ramos, M.C., Branch, C.L., Pitera, A.M., Kozlovsky, D.Y., Bridge, E.S., and Pravosudov, V.V. (2018). Memory in wild mountain chickadees from different elevations: comparing first year birds with older survivors. Anim. Behav. *137*, 149–160.
16. Heinen, V.K., Pitera, A.M., Sonnenberg, B.R., Benedict, L.M., Branch, C.L., Bridge, E.S., and Pravosudov, V.V. (2021). Specialized spatial cognition is associated with reduced cognitive senescence in a food-caching bird. Proc. Biol. Sci. *288*, 20203180.
17. Croston, R., Branch, C.L., Kozlovsky, D.Y., Dukas, R., and Pravosudov, V.V. (2015). Heritability and the evolution of cognitive traits. Behav. Ecol. *26*, 1447–1459.
18. Endler, J.A. (1986). Natural Selection in the Wild (Princeton University Press).
19. Branch, C.L., Jahner, J.P., Kozlovsky, D.Y., Parchman, T.L., and Pravosudov, V.V. (2017). Absence of population structure across elevational gradients despite large phenotypic variation in mountain chickadees (*Poecile gambeli*). R. Soc. Open Sci. *4*, 170057.
20. Brieuc, M.S.O., Waters, C.D., Drinan, D.P., and Naish, K.A. (2018). A practical introduction to Random Forest for genetic association studies in ecology and evolution. Mol. Ecol. Resour. *18*, 755–766.
21. Croston, R., Branch, C.L., Pitera, A.M., Kozlovsky, D.Y., Bridge, E.S., Parchman, T.L., and Pravosudov, V.V. (2017). Predictably harsh environment is associated with reduced cognitive flexibility in wild food-caching mountain chickadees. Anim. Behav. *123*, 139–149.
22. Branch, C.L., Pitera, A.M., Kozlovsky, D.Y., Bridge, E.S., and Pravosudov, V.V. (2019). Smart is the new sexy: female mountain chickadees increase reproductive investment when mated to males with better spatial cognition. Ecol. Lett. *22*, 897–903.
23. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.
24. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. *9*, e1003264.
25. Freda, P.J., Alex, J.T., Morgan, T.J., and Ragland, G.J. (2017). Genetic decoupling of thermal hardiness across metamorphosis in Drosophila melanogaster. Integr. Comp. Biol. *57*, 999–1009.
26. Semenov, G.A., Linck, E., Enbody, E.D., Harris, R.B., Khaydarov, D.R., Alström, P., Andersson, L., and Taylor, S.A. (2021). Asymmetric introgression reveals the genetic architecture of a plumage trait. Nat. Commun. *12*, 1019.
27. Díez-Zaera, M., Díaz-Hernández, J.I., Hernández-Álvarez, E., Zimmermann, H., Díaz-Hernández, M., and Miras-Portugal, M.T. (2011). Tissue-nonspecific alkaline phosphatase promotes axonal growth of hippocampal neurons. Mol. Biol. Cell *22*, 1014–1024.
28. Gonda, Y., Namba, T., and Hanashima, C. (2020). Beyond axon guidance: roles of Slit-Robo signaling in neocortical formation. Front. Cell Dev. Biol. *8*, 607415.
29. Roelink, H. (2000). Hippocampus formation: an intriguing collaboration. Curr. Biol. *10*, R279–R281.
30. Epifanova, E., Babaev, A., Newman, A.G., and Tarabykin, V. (2019). Role of Zeb2/Sip1 in neuronal development. Brain Res. *1705*, 24–31.
31. Yang, Q.Q., Zhai, Y.Q., Wang, H.F., Cai, Y.C., Ma, X.Y., Yin, Y.Q., Li, Y.D., Zhou, G.M., Zhang, X., Hu, G., and Zhou, J.W. (2021). Nuclear isoform of FGF13 regulates post-natal neurogenesis in the hippocampus through an epigenomic mechanism. Cell Rep. *35*, 109127.
32. Zuko, A., Oguro-Ando, A., van Dijk, R., Gregorio-Jordan, S., van der Zwaag, B., and Burbach, J.P.H. (2016). Developmental role of the cell adhesion molecule Contactin-6 in the cerebral cortex and hippocampus. Cell Adhes. Migr. *10*, 378–392.
33. Cordner, R.D., Friend, L.N., Mayo, J.L., Badgley, C., Wallmann, A., Stallings, C.N., Young, P.L., Miles, D.R., Edwards, J.G., and Bridgewater, L.C. (2017). The BMP2 nuclear variant, nBMP2, is expressed in mouse hippocampus and impacts memory. Sci. Rep. *7*, 46464.
34. Oku, Y., and Huganir, R.L. (2013). AGAP3 and Arf6 regulate trafficking of AMPA receptors and synaptic plasticity. J. Neurosci. *33*, 12586–12598.
35. Herold, C. (2010). NMDA and D2-like receptors modulate cognitive flexibility in a color discrimination reversal task in pigeons. Behav. Neurosci. *124*, 381–390.
36. Audet, J.-N., Kayello, L., Ducatez, S., Perillo, S., Cauchard, L., Howard, J.T., O'Connell, L.A., Jarvis, E.D., and Lefebvre, L. (2018). Divergence

in problem-solving skills is associated with differential expression of glutamate receptors in wild finches. Sci. Adv. *4*, eaao6369.

37. Sun, C., Qiu, Y., Ren, Q., Zhang, X., Cao, B., Zou, Y., Li, J., Zhang, J., and Wang, Y. (2021). Molecular cloning and functional characterization of three 5-HT receptor genes (HTR1B, HTR1E, and HTR1F) in chickens. Genes (Basel) *12*, 891.
38. Sun, Y., Zhang, F., Gao, J., Gao, X., Guo, T., Zhang, K., Shi, Y., Zheng, Z., Tang, W., Zheng, Y., et al. (2006). Positive association between POU1F1 and mental retardation in young females in the Chinese Han population. Hum. Mol. Genet. *15*, 1237–1243.
39. Stokholm, J., Teasdale, T.W., Johannsen, P., Nielsen, J.E., Nielsen, T.T., Isaacs, A., Brown, J.M., and Gade, A.; Frontotemporal dementia Research in Jutland Association (FReJA) consortium (2013). Cognitive impairment in the preclinical stage of dementia in FTD-3 CHMP2B mutation carriers: a longitudinal prospective study. J. Neurol. Neurosurg. Psychiatry *84*, 170–176.
40. Nyberg, F., and Hallberg, M. (2013). Growth hormone and cognitive function. Nat. Rev. Endocrinol. *9*, 357–365.
41. Studzinski, A.L.M., Barros, D.M., and Marins, L.F. (2015). Growth hormone (GH) increases cognition and expression of ionotropic glutamate receptors (AMPA and NMDA) in transgenic zebrafish (Danio rerio). Behav. Brain Res. *294*, 36–42.
42. Cuveillier, C., Boulan, B., Ravanello, C., Denarier, E., Deloulme, J.-C., Gory-Fauré, S., Delphin, C., Bosc, C., Arnal, I., and Andrieux, A. (2021). Beyond neuronal microtubule stabilization: MAP6 and CRMPS, two converging stories. Front. Mol. Neurosci. *14*, 665693.
43. Fournet, V., Schweitzer, A., Chevarin, C., Deloulme, J.-C., Hamon, M., Giros, B., Andrieux, A., and Martres, M.-P. (2012). The deletion of STOP/MAP6 protein in mice triggers highly altered mood and impaired cognitive performances. J. Neurochem. *121*, 99–114.
44. Sun, L.-L., Yang, S.-L., Sun, H., Li, W.-D., and Duan, S.-R. (2019). Molecular differences in Alzheimer's disease between male and female patients determined by integrative network analysis. J. Cell. Mol. Med. *23*, 47–58.
45. Datson, N.A., Polman, J.A.E., de Jonge, R.T., van Boheemen, P.T.M., van Maanen, E.M.T., Welten, J., McEwen, B.S., Meiland, H.C., and Meijer, O.C. (2011). Specific regulatory motifs predict glucocorticoid responsiveness of hippocampal gene expression. Endocrinology *152*, 3749–3757.
46. Schwartz, F., Neve, R., Eisenman, R., Gessler, M., and Bruns, G. (1994). A WAGR region gene between PAX-6 and FSHB expressed in fetal brain. Hum. Genet. *94*, 658–664.
47. Ohadi, M., Mirabzadeh, A., Esmaeilzadeh-Gharehdaghi, E., Rezazadeh, M., Hosseinkhanni, S., Oladnabi, M., Firouzabadi, S.G., and Darvish, H. (2012). Novel evidence of the involvement of calreticulin in major psychiatric disorders. Prog. Neuropsychopharmacol. Biol. Psychiatry *37*, 276–281.
48. Fragopoulou, A.F., Polyzos, A., Papadopoulou, M.-D., Sansone, A., Manta, A.K., Balafas, E., Kostomitsopoulos, N., Skouroliakou, A., Chatgilialoglu, C., Georgakilas, A., et al. (2018). Hippocampal lipidome and transcriptome profile alterations triggered by acute exposure of mice to GSM 1800 MHz mobile phone radiation: an exploratory study. Brain Behav. *8*, e01001.
49. Reggiani, C., Coppens, S., Sekhara, T., Dimov, I., Pichon, B., Lufin, N., Addor, M.-C., Belligni, E.F., Digilio, M.C., Faletra, F., et al. (2017). Novel promoters and coding first exons in DLG2 linked to developmental disorders and intellectual disability. Genome Med. *9*, 67.
50. Nesan, D., and Vijayan, M.M. (2013). The transcriptomics of glucocorticoid receptor signaling in developing zebrafish. PLoS ONE *8*, e80726.
51. Zhao, B., Shan, Y., Yang, Y., Yu, Z., Li, T., Wang, X., Luo, T., Zhu, Z., Sullivan, P., Zhao, H., et al. (2021). Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. Nat. Commun. *12*, 2878.
52. Yang, X.Y., Stanley, R.E., Ross, A.P., Robitaille, A.M., Gray, J.A., and Cheyette, B.N.R. (2019). Sestd1 encodes a developmentally dynamic synapse protein that complexes with BCR Rac1-GAP to regulate forebrain dendrite, spine and synapse formation. Cereb. Cortex *29*, 505–516.
53. Jang, S., Oh, D., Lee, Y., Hosy, E., Shin, H., van Riesen, C., Whitcomb, D., Warburton, J.M., Jo, J., Kim, D., et al. (2016). Synaptic adhesion molecule IgSF11 regulates synaptic transmission and plasticity. Nat. Neurosci. *19*, 84–93.
54. Lo Bianco, L., Attrotto, M.T., Torretta, S., Masellis, R., Rampino, A., D'Ambrosio, E., Di Giorgio, A., Ferranti, L., Fazio, L., Gelao, B., et al. (2017). Genetic variation is associated with RTN4R expression and working memory processing in healthy humans. Brain Res. Bull. *134*, 162–167.
55. Theocharidis, U., Roll, L., and Faissner, A. (2021). The expression of tenascin-C in neural stem/progenitor cells is stimulated by the growth factors EGF and FGF-2, but not by TGFβ1. Cell Tissue Res. *385*, 659–674.
56. Sun, D., Tan, Z.-B., Sun, X.D., Liu, Z.P., Chen, W.B., Milibari, L., Ren, X., Yao, L.L., Lee, D., Shen, C., et al. (2021). Hippocampal astrocytic neogenin regulating glutamate uptake, a critical pathway for preventing epileptic response. Proc. Natl. Acad. Sci. USA *118*, e2022921118.
57. Wirthlin, M., Lima, N.C.B., Guedes, R.L.M., Soares, A.E.R., Almeida, L.G.P., Cavaleiro, N.P., Loss de Morais, G., Chaves, A.V., Howard, J.T., Teixeira, M.M., et al. (2018). Parrot genomes and the evolution of heightened longevity and cognition. Curr. Biol. *28*, 4001–4008.e7.
58. van der Ende, E.L., Meeter, L.H., Stingl, C., van Rooij, J.G.J., Stoop, M.P., Nijholt, D.A.T., Sanchez-Valle, R., Graff, C., Öijerstedt, L., Grossman, M., et al. (2019). Novel CSF biomarkers in genetic frontotemporal dementia identified by proteomics. Ann. Clin. Transl. Neurol. *6*, 698–707.
59. Lanke, V., Moolamalla, S.T.R., Roy, D., and Vinod, P.K. (2018). Integrative analysis of hippocampus gene expression profiles identifies network alterations in aging and Alzheimer's disease. Front. Aging Neurosci. *10*, 153.
60. Tschentscher, A., Dekomien, G., Ross, S., Cremer, K., Kukuk, G.M., Epplen, J.T., and Hoffjan, S. (2015). Analysis of the C19orf12 and WDR45 genes in patients with neurodegeneration with brain iron accumulation. J. Neurol. Sci. *349*, 105–109.
61. Wang, Z., Andrade, N., Torp, M., Wattananit, S., Arvidsson, A., Kokaia, Z., Jørgensen, J.R., and Lindvall, O. (2012). Meteorin is a chemokinetic factor in neuroblast migration and promotes stroke-induced striatal neurogenesis. J. Cereb. Blood Flow Metab. *32*, 387–398.
62. Cai, W., Wei, Y., Jarnik, M., Reich, J., and Lilly, M.A. (2016). The GATOR2 component Wdr24 regulates TORC1 activity and lysosome function. PLoS Genet. *12*, e1006036.
63. Lastun, V.L., Grieve, A.G., and Freeman, M. (2016). Substrates and physiological functions of secretase rhomboid proteases. Semin. Cell Dev. Biol. *60*, 10–18.
64. Uddin, M.G., Siddiqui, S.A., Uddin, M.S., Aziz, M.A., Hussain, M.S., Furhatun-Noor, M.S., Sen Millat, N., Muhiri, B., and Islam, M.S. (2020). Genetic variants of ZNF385B and COMT are associated with autism spectrum disorder in the Bangladeshi children. Meta Gene *26*, 100820.
65. Dumas, G., Malesys, S., and Bourgeron, T. (2021). Systematic detection of brain protein-coding genes under positive selection during primate evolution and their roles in cognition. Genome Res. *31*, 484–496.
66. Nielsen, J.V., Thomassen, M., Møllgård, K., Noraberg, J., and Jensen, N.A. (2014). Zbtb20 defines a hippocampal neuronal identity through direct repression of genes that control projection neuron development in the isocortex. Cereb. Cortex *24*, 1216–1229.
67. Francis, H.M., Mirzaei, M., Pardey, M.C., Haynes, P.A., and Cornish, J.L. (2013). Proteomic analysis of the dorsal and ventral hippocampus of rats maintained on a high fat and refined sugar diet. Proteomics *13*, 3076–3091.
68. Sahún, I., Marechal, D., Pereira, P.L., Nalesso, V., Gruart, A., Garcia, J.M.D., Antonarakis, S.E., Dierssen, M., and Herault, Y. (2014). Cognition and hippocampal plasticity in the mouse is altered by

monosomy of a genomic region implicated in Down syndrome. Genetics *197*, 899–912.

69. Díaz-Morán, S., Palència, M., Mont-Cardona, C., Cañete, T., Blázquez, G., Martínez-Membrives, E., López-Aumatell, R., Sabariego, M., Donaire, R., Morón, I., et al. (2013). Gene expression in hippocampus as a function of differential trait anxiety levels in genetically heterogeneous NIH-HS rats. Behav. Brain Res. *257*, 129–139.
70. Yeo, K.S., Tan, M.C., Wong, W.Y., Loh, S.W., Lam, Y.L., Tan, C.L., Lim, Y.-Y., and Ea, C.-K. (2016). JMJD8 is a positive regulator of TNF-induced NF-κB signaling. Sci. Rep. *6*, 34125.
71. Buschdorf, J.P., Ong, M.L., Ong, S.X., MacIsaac, J.L., Chng, K., Kobor, M.S., Meaney, M.J., and Holbrook, J.D. (2016). Low birth weight associates with hippocampal gene expression. Neuroscience *318*, 190–205.
72. Mir, Y.R., and Kuchay, R.A.H. (2019). Advances in identification of genes involved in autosomal recessive intellectual disability: a brief review. J. Med. Genet. *56*, 567–573.
73. Li, S., Papale, L.A., Zhang, Q., Madrid, A., Chen, L., Chopra, P., Keleş, S., Jin, P., and Alisch, R.S. (2016). Genome-wide alterations in hippocampal 5-hydroxymethylcytosine links plasticity genes to acute stress. Neurobiol. Dis. *86*, 99–108.
74. Li, K., Müller, I., Patil, S., Höger, H., Pollak, A., Russo-Schlaff, N., Lubec, G., and Li, L. (2012). Strain-independent global effect of hippocampal proteins in mice trained in the Morris water maze. Amino Acids *43*, 1739–1749.
75. Li, Q., Wang, Z., Zong, L., Ye, L., Ye, J., Ou, H., Jiang, T., Guo, B., Yang, Q., Liang, W., et al. (2021). Allele-specific DNA methylation maps in monozygotic twins discordant for psychiatric disorders reveal that disease-associated switching at the EIPR1 regulatory loci modulates neural function. Mol. Psychiatry. Published online May 7, 2021. https://doi.org/10.1038/s41380-021-01126-w.
76. Sanchez, D., Bajo-Grañeras, R., Del Caño-Espinel, M., Garcia-Centeno, R., Garcia-Mateo, N., Pascua-Maestro, R., and Ganfornina, M.D. (2015). Aging without Apolipoprotein D: molecular and cellular modifications in the hippocampus and cortex. Exp. Gerontol. *67*, 19–47.
77. Egan, M.F., Straub, R.E., Goldberg, T.E., Yakub, I., Callicott, J.H., Hariri, A.R., Mattay, V.S., Bertolino, A., Hyde, T.M., Shannon-Weickert, C., et al. (2004). Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. Proc. Natl. Acad. Sci. USA *101*, 12604–12609.
78. Li, W., Tam, K.M.V., Chan, W.W.R., Koon, A.C., Ngo, J.C.K., Chan, H.Y.E., and Lau, K.-F. (2018). Neuronal adaptor FE65 stimulates Rac1-mediated neurite outgrowth by recruiting and activating ELMO1. J. Biol. Chem. *293*, 7674–7688.
79. Lu, Z., Elliott, M.R., Chen, Y., Walsh, J.T., Klibanov, A.L., Ravichandran, K.S., and Kipnis, J. (2011). Phagocytic activity of neuronal progenitors regulates adult neurogenesis. Nat. Cell Biol. *13*, 1076–1083.
80. Kempermann, G., and Gage, F.H. (1999). New nerve cells for the adult brain. Sci. Am. *280*, 48–53.
81. Ming, G.L., and Song, H. (2005). Adult neurogenesis in the mammalian central nervous system. Annu. Rev. Neurosci. *28*, 223–250.
82. Kempermann, G., Wiskott, L., and Gage, F.H. (2004). Functional significance of adult neurogenesis. Curr. Opin. Neurobiol. *14*, 186–191.
83. Barnea, A., and Pravosudov, V. (2011). Birds as a model to study adult neurogenesis: bridging evolutionary, comparative and neuroethological approaches. Eur. J. Neurosci. *34*, 884–907.
84. Zador, A.M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. Nat. Commun. *10*, 3770.
85. Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749–1751.
86. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). Grp Genomes Project Anal, the variant call format and VCFtools. J. Bioinform. *27*, 2156–2158.
87. Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2*, 18–22.
88. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and 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 (Austin) *6*, 80–92.
89. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. J. Bioinform. *37*, 1639–1643.
90. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
91. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data, 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
92. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45* (D1), D183–D189.
93. Branch, C.L., Pitera, A.M., Kozlovsky, D.Y., Sonnenberg, B., Benedict, L., and Pravosudov, V.V. (2019). Elevation-related differences in the age structure of breeding birds suggest stronger selection at harsher elevations. Behav. Ecol. Sociobiol. *73*, 143.
94. Branch, C.L., Sonnenberg, B.R., Pitera, A.M., Benedict, L.M., Kozlovsky, D.Y., Bridge, E.S., and Pravosudov, V.V. (2020). Testing the greater male variability phenomenon: male mountain chickadees exhibit larger variation in reversal learning performance compared with females. Proc. Biol. Sci. *287*, 20200895.
95. Branch, C.L., Kozlovsky, D.Y., and Pravosudov, V.V. (2015). Elevation related variation in aggressive response to mirror image in mountain chickadees. Behaviour *152*, 667–676.
96. Kozlovsky, D.Y., Branch, C.L., Freas, C.A., and Pravosudov, V.V. (2014). Elevation related differences in exploration and social dominance in mountain chickadees (*Poecile gambeli*). Behav. Ecol. Sociobiol. *68*, 1871–1881.
97. Branch, C.L., and Pravosudov, V.V. (2015). Mountain chickadees from different elevations sing different songs: acoustic adaptation, temporal drift or signal of local adaptation? R. Soc. Open Sci. *2*, 150019.
98. Pitera, A.M., Branch, C.L., Bridge, E.S., and Pravosudov, V.V. (2018). Daily foraging routines in food-caching mountain chickadees are associated with variation in environmental harshness. Anim. Behav. *143*, 93–104.
99. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
100. Broad Institute (2019). Picard Tools (Broad Institute). http://broadinstitute.github.io/picard.
101. Bates, D., Marchler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48.
102. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: tests in linear mixed effects models. J. Stat. Softw. *82*, 1–26.
103. Cirulli, E.T., and Goldstein, D.B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. *11*, 415–425.
104. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
105. Toews, D.P., Taylor, S.A., Vallender, R., Brelsford, A., Butcher, B.G., Messer, P.W., and Lovette, I.J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. *26*, 2313–2318.

106. Gazda, M.A., Araújo, P.M., Lopes, R.J., Toomey, M.B., Andrade, P., Afonso, S., Marques, C., Nunes, L., Pereira, P., Trigo, S., et al. (2020). A genetic mechanism for sexual dichromatism in birds. Science *368*, 1270–1274.

107. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

108. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

109. Zhang, G., Karns, R., Sun, G., Indugula, S.R., Cheng, H., Havas-Augustin, D., Novokmet, N., Durakovic, Z., Missoni, S., Chakraborty, R., et al. (2012). Finding missing heritability in less significant Loci and allelic heterogeneity: genetic variation in human height. PLoS ONE *7*, e51211.

110. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.

111. Wei, W.-H., Hemani, G., and Haley, C.S. (2014). Detecting epistasis in human complex traits. Nat. Rev. Genet. *15*, 722–733.

112. Basu, S., Kumbier, K., Brown, J.B., and Yu, B. (2018). Iterative random forests to discover predictive and stable high-order interactions. Proc. Natl. Acad. Sci. USA *115*, 1943–1948.

113. Goldstein, B.A., Hubbard, A.E., Cutler, A., and Barcellos, L.F. (2010). An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. BMC Genet. *11*, 49.

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Blood samples from wild birds	Mountain chickadees	N/A
Critical commercial assays		
DNeasy Blood and Tissue DNA extraction kit	QIAGEN	69506
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
Deposited data		
Raw cognitive testing data	This paper	Data S1
Black-capped chickadee reference genome	[12]	NCBI: JAAMOC000000000
Raw paired whole genome sequencing reads	This paper	NCBI: PRJNA770082
Experimental models: Organisms/strains		
Mountain chickadees (bird)	Wild caught	*Poecile gambeli*
Software and algorithms		
GATK	Broad Institute	ver. 4.1.0.0
Picard Tools	Broad Institute	ver. 2.22.7
bcftools	[85]	ver. 1.7
vcftools	[86]	ver. 0.1.13
GEMMA	[23]	ver. 0.98
randomForest (R package)	[87]	ver. 4.6-14
SnpEff	[88]	ver. 4.3
Lifoff	[89]	ver. 1.5.2
R	R Core Team	ver. 3.6.3
Trimmomatic	[90]	ver. 0.39
FastQC	[91]	ver. 0.11.6
PANTHER	[92]	ver. 15

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Vladimir Pravosudov (vpravosu@unr.edu).

Materials availability

This study did not generate new unique reagents

Data and code availability

Raw cognitive testing data for this study is available as supplemental data file included with this manuscript (Data S1). Raw paired whole genome sequencing reads are available through NCBI: PRJNA770082.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study subjects and site

We selected individual birds in this study based on their performance on a spatial learning and memory task at our long-term study site in northern California, Sagehen Experimental Forest, USA.[7,9,14–16,19,22] Birds were tested in the wild using 'smart' feeder spatial

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

arrays equipped with radiofrequency identification (RFID) technology (see diagram in Figure 1D and Video S1).[9,14–16,21,22] Prior to each year of testing, chickadees were trapped using mist nets at established feeders and their legs were fitted with unique color-band and passive integrated transponder (PIT)-tag combinations. Birds were selected from 3 years of testing (2016 – 2018) based on the mean number of errors they made across the entire testing period (16 days in 2016 and 5 days in 2017 and 2018; see more detailed methods in 9 and 14). Birds from both high (ca 2400 m) and low (ca 1900 m) elevations were selected as best (few errors) or worst (many errors) performers for subsequent genetic analyses. Birds were chosen as best and worst performers from the bottom and top of the data distributions (Figure 1D), respectively, and selected such that all best performers were detected in more than one year (22/22) and most of the worst performers were not detected in more than one year (15/20: 9 of 10 birds at high elevation and 6 of 10 birds at low elevation, Differences between best and worst performers: Fisher's exact test, $p < 0.001$; e.g., 13). Due to the sedentary nature of chickadees, birds that were not detected in the following year after testing were presumed to have died.[93] We aimed for a total sample size of 40 birds, which meant we selected the 10 best and 10 worst performers from high elevation and the 10 best and 10 worst performers from low elevation. Although we refer to birds in best and worst performance groups, the nature of our selection process (and likely, the polygenic architecture of cognition) resulted in a continuous distribution of individual performance scores used for all genetic analyses described throughout the manuscript (Figure 1F). Male and female chickadees in this population do not differ in performance on the spatial cognitive task,[94] however we did use sex as a covariate in GWAS (see below).

METHOD DETAILS

Cognitive testing in the wild

Birds were tested on the spatial cognitive task using 4 spatial arrays (example of one array in Figure 1D; 2 at each elevation), each containing 8 RFID-enabled feeders mounted equidistant on an aluminum square frame (1.2 × 1.2 m), suspended 2-3 m above the ground using a pulley system connected to 4 trees (to avoid damage by squirrels and black bears). Within each elevation, the two arrays were positioned ca. 1.5 km apart and each was visited by mostly non-overlapping groups of chickadees. Each feeder has a perch with an embedded RFID antenna that is mounted in front of a motorized door that allows access to a black oil sunflower seed reward (Video S1). Feeders can function in one of three modes: (a) 'open' mode, where the door remains open with visible food; (b) 'all' mode, where the door remains closed but opens when any PIT-tagged bird lands on the perch, allowing access to food; and (c) 'target' mode, where the door opens only for PIT-tag IDs that have been programmed into the RFID reader memory. 'Open' and 'all' modes are largely used for training whereas 'target' mode allows us to restrict food access for individual birds to a specific feeder. In all three modes, every feeder records the PIT-tag ID, date, and time of all visits. For each testing year, birds were habituated to the moving feeder door (i.e., 'all' mode) for at least 2 weeks prior to testing. Birds that were consistently visiting the feeder arrays were pseudo-randomly assigned to one individual feeder within the array (1 of 8). The measure of spatial cognitive ability was assessed based on a birds' visits to other feeders within the array (7 of 8) that were not its assigned rewarding feeder. When a bird landed on the perch (with embedded RFID antenna) of its assigned feeder, the feeder door would open, and the bird would obtain one sunflower seed. Unlike many fringillid or corvid species, chickadees do not consume multiple seeds or monopolize a feeder, but instead take one seed and either fly to a nearby tree to consume it or fly further to cache it. Therefore, we are confident that birds received one seed at each correct feeder visit. Performance on the spatial task was measured as the number of location errors an individual made within a trial. A trial began when the bird visited any feeder within the array and ended when they visited their rewarding feeder, at which time the number of location errors was reset to zero and a new trial started. Location errors were defined as the number of unrewarding feeders a bird visited before landing on the correct, rewarding feeder. The mean number of location errors per trial was calculated across the entire testing period and used to choose birds as best or worst based on their performance.

For this study we chose birds with particular cognitive phenotypes to assess genetic variation among best and worst performers. It is worth noting that our previous work on cognitive variation in caching birds has focused on birds inhabiting locations that differ in winter climate severity (e.g., high versus low elevations).[9,14,15,21] In addition to cognitive variation at the group level, these birds also vary in other traits, including aggression,[95] social dominance, novel exploration,[96] song,[97] and daily foraging routines.[98] While we cannot completely rule out the possibility of confounding variation among the best and worst performing birds used in this study, including birds from both high and low elevations in our best and worst performer groups aimed to reduce spurious correlations (Figure 1E).

Library preparation and whole genome sequencing

We extracted DNA from blood stored in Queens lysis buffer using the QIAGEN DNeasy Blood and Tissue protocol and quantified it using an Invitrogen Qubit 3.0 fluorometer (Invitrogen, Carlsbad CA) and the double-stranded DNA broad range assay kit. Whole genome library preparation was carried out by the University of Colorado Boulder Next Generation Sequencing Facility using a Nextera XT DNA Library Preparation Kit following standard protocol, except using half reaction volumes. Libraries were pooled and sequenced on an Illumina NovaSEQ 6000 (paired end, 150 base pairs) at the University of Colorado Cancer Center Genomics and Microarray Core Facility. Sequencing resulted in a total of 1 billion reads across all individuals, representing an average of 23 million reads per individual (17,124,662 – 73,759,522), and an average depth of coverage of 9.43X per individual. Filtering of raw reads resulted in 671 Gb of data among 42 individuals. We trimmed raw reads of sequencing adapters and barcodes using Trimmomatic (ver. 0.39)[90] using the paired end (PE) settings and Illumina adapters. We performed quality control on trimmed sequence files using FastQC (ver. 0.11.6).[91] All sequences passed FastQC quality control.

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

Whole genome alignment and filtering

We aligned filtered sequence data to the black-capped chickadee reference genome (https://www.ncbi.nlm.nih.gov/bioproject/589043; Accession: JAAMOC000000000)[12] according to GATK "Best Practices" using the GATK software (ver. 4.1.0.0).[99] We used the black-capped chickadee reference genome because no reference genome is available for the mountain chickadee at the moment and these species are closely related. After initial alignment with Picard Tools (ver. 2.22.7),[100] we used "mpileup" from the bcftools software (ver. 1.7)[85] to identify single nucleotide polymorphisms (SNPs) and generate a variant call file (VCF). Alignment of filtered sequence data before VCF table filtering initially yielded 41,105,895 SNPs. Using vcftools (ver. 0.1.13),[86] we initially removed any genotype with a sequencing depth of less than 5, a quality score lower than 30, or a minor allele count less than 3, and checked that no individuals had genotype missingness greater than 10% within each group. We removed any locus that did not pass a Hardy-Weinberg Equilibrium test (P-value < 0.05) and removed all indels. We filtered SNPs so that all individuals had at least 95% of all genotypes and a minor allele frequency of 10%. Data filtering yielded 12,106,779 SNPs in 42 individuals (20 individuals who performed the worst on spatial cognition tests, the most of which did not survive past their first year, and 22 individuals who performed the best on spatial cognition tests and survived beyond their first year). There were 24 males and 18 females in the dataset, but they were distributed relatively equally across performance groups (12:8 males/females in the low performance group and 12:10 males/females in the high performance group; furthermore, there are no significant differences in spatial cognitive performance between males and females in this population[94]).

QUANTIFICATION AND STATISTICAL ANALYSIS

Measuring cognitive ability

We assessed the mean number of errors per trial over the entire testing period. This same measure has been used in our previous studies on spatial cognition in wild mountain chickadees[9,14,15,21] and its ecological relevance has been previously established, showing that spatial cognitive ability predicts overwinter survival[14] and that females increase reproductive output when mated to males with better spatial cognition.[22] Spatial cognitive performance of best and worst individuals was compared using general linear mixed models (GLMM; *lmer()* function in R version 3.6.1, packages lme4 and lmerTest[101,102]) with mean performance across all trials as the response variable; cognitive category (2 levels: best and worst), elevation (2 levels: high and low), and total number of trials completed as fixed factors, and testing year as a random factor (see Data S1). Total number of trials was used as a fixed factor to control for the effect of motivation and differences in the number of trials completed across birds over the fixed testing period.

Regression GWAS analysis

For GWAS, we combined individuals from high and low elevation sites because of a lack of population structure between the sites and banding-based records demonstrating effective migration along the elevational gradient.[19] We used GEMMA v.0.98 (Genome-wide Efficient Mixed-Model Association[23]) to perform Linear Mixed-effects Models (LMM) to establish associations between cognition phenotype (total number of errors, see above) and individual SNPs (complete dataset of 12,106,779 loci). To account for potential relatedness between individuals and potential effects of population stratification (e.g., recent migrants from other populations), we estimated the relatedness matrix (gk −1 option of GEMMA) and supplied it as a covariate along with binary sex information in our LMM analyses. We used Wald likelihood ratio test to assess significance of association under P value of 1.00E-05 (-log10(P) = 5) and 1.00E-06 (-log10(P) = 6). These thresholds were chosen to minimize the number of expected false positives given the number of loci in our dataset (120 false positives are expected for the former and 12 for the latter thresholds), while maximizing the number of associations to use for gene ontology analyses. We used Bayesian Sparse Linear Mixed Models (BSLMM)[24] implemented in GEMMA to assess the genetic architecture of cognition. We ran four MCMC chains with five million steps and a subsequent 20 million MCMC step sampling every 1000 iterations. We assessed three BSLMM hyperparameters: PVE (the proportion of variance explained by all SNPs), PGE (the proportion of genetic variance explained by alleles with measurable effect), and posterior number of SNPs explaining trait variance in the model. Finally, we used a leave-one-out cross-validation approach to assess the posterior predictive power of BSLMM (Figure S1). More specifically, we excluded phenotypic information for one individual at a time and used the remaining dataset to predict its phenotype based on genotype by running BSLMM with the settings indicated above. Linear regression was used to estimate the proportion of variance explained by predicted phenotypes as a measure of predictive performance.

Random Forest filtering and analysis

Advances in genome sequencing and analyses have improved our ability to identify genes that underlie phenotypes when they have a simple genetic basis, resulting in important insights into a variety of biological processes, from human disease[103,104] to plumage coloration[105] and sexual dichromatism[106] in birds. Traditionally, regression-based genome wide association studies (GWAS) have been used for such investigations; however, these analyses are sometimes underpowered for detecting genotype associations with polygenic phenotypes because of the statistical difficulty of detecting multiple genes each with a small effect on a phenotype (e.g., human height[107–111]). In addition, most GWAS algorithms assume co-dominant and additive inheritance models of phenotypic traits and are not well-suited for detecting complete or partial dominance and epistatic interactions between loci.[112] Machine learning is a promising avenue for overcoming these issues and can provide valuable information complementary to GWAS.[20,87]

We performed a Random Forest analysis using the same 42 individuals with the best and the worst spatial cognitive performance that were used for GWAS. We excluded loci that did not have 100% of genotypes in all individuals using vcftools. To make our dataset

Please cite this article in press as: Branch et al., The genetic basis of spatial cognitive variation in a food-caching bird, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.10.036

computationally manageable for Random Forest, we further thinned it to retain one SNP per 500bp window using vcftools, resulting in 1,312,917 SNPs. We performed a regression Random Forest analysis using the R package "randomForest" (ver. 4.6-14)[113] in relation to the spatial cognition phenotype, closely following the analysis outline and code from 19. The two parameters that have the most influence on the OOB-ER or PVE are the number of trees grown (num.trees) per forest and the number of predictors to randomly sample at each node (mtry).[113] We first ran the parameter optimization while simultaneously testing a range of num.trees (from 100 to 13,100 with a step of 500) and mtry: sqrt(n.loci), 2sqrt(n.loci), 20sqrt(n.loci), 0.1 n.loci, n.loci /4, n.loci /2. This analysis revealed that num.trees values reached a plateau at ~2,000 and that mtry of ~23,000 maximized the PVE. Following recommendations, we increased num.trees of trees by five times (10,000) and used mtry of 23,000 for the subsequent analysis.[113] We ran three independent Random Forests to test for correlation between importance scores assigned to individual loci. The correlation between all loci was very low (r^2 not exceeding 0.07, also tested with a broader range of num.trees and mtry) which was attributed to the majority of loci having an unstable association signal, such as slightly negative importance values assigned in one analysis replicate, slightly positive in the other and zero in the third. We subsampled the top 2 to 30% of loci with highest importance values and ran three separate Random Forests on these datasets to identify the breakpoint of r^2 increase. This revealed that a small fraction of loci have the strongest association signal (best2% r^2 = 0.27, best3% r^2 = 0.27, best4% r^2 = 0.1, best5% r^2 = 0.1, best10% r^2 = 0.1, best20% r^2 = 0.1, best30% r^2 = 0.1), with additional analyses indicating the breakpoint at around top 0.1% (n = 1,312), where r^2 increased to 0.75. We therefore used the top 0.1% of outliers for downstream analyses.

Gene ontology analysis

To identify which biological processes may be represented by the SNPs identified as significantly related to the spatial cognition phenotype in the regression GWAS analysis and the Random Forest analysis, we first annotated the SNPs using SnpEff (ver. 4.3t)[88] with the black-capped chickadee reference genome.[12] We identified unique, known genes (i.e., those with a defined gene name and accession number) to be used in the gene ontology analysis. After genes were annotated, we compared gene lists derived from SNPs identified as important in both the regression GWAS analysis and the Random Forest analysis.

For the gene ontology analysis of the regression GWAS, we used all genes with a *P*-value < 1.00E-05. For the Random Forest analysis, we included all genes with 0.1% top positive importance score as indicated above. We used these cutoffs in order to maximize the number of genes used in the gene ontology analysis, because we expect many genes with potentially small individual effects to contribute to a complex phenotype, such as spatial cognition. We used PANTHER (ver. 15)[92] to perform the gene ontology analysis. Specifically, we tested for statistical overrepresentation in biological processes using the "PANTHER GO-Biological Process Complete" analysis using the Bonferroni correction for multiple testing. We performed the analyses using the human (*Homo sapiens*) and the chicken (*Gallus gallus*) reference genomes. We considered both reference genomes because the human genome has a more complete annotation (and better characterized genes involved in behavior and cognition), while the chicken genome represents a more closely related species to chickadees (Tables S1 and S3).

Gene annotation

To examine associations between Random Forest and GEMMA outliers and specific genes, we used Liftoff[89] to annotate our black-capped chickadee reference genome, using zebra finch mRNA and protein evidence (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/008/822/105/GCF_008822105.2_bTaeGut2.pat.W.v2/). This annotation resulted in 19,482 out of 21,049 (~93%) of zebra finch genes mapped to the black-capped chickadee reference. We then used a 200,000bp window around specific SNPs to examine gene content and associations between outliers and individual gene features (such as coding versus non-coding regions, intronic versus exonic parts of a gene, etc). Given that our GEMMA analysis of the cognition phenotype architecture revealed that ~10 loci might be explaining a high proportion of the trait variance, we focused on 14 genomic regions with the exceptional outlier values of GEMMA, Random Forest or both (Figures 3 and S3).