

# Epigenetic Divergence during Early Stages of Speciation in an African Crater Lake Cichlid Fish

Grégoire Vernaz<sup>1,2,3\*</sup>, Alan G. Hudson<sup>4,8</sup>, M. Emília Santos<sup>5</sup>, Bettina Fischer<sup>2</sup>, Madeleine Carruthers<sup>4</sup>, Asilatu H. Shechonge<sup>6</sup>, Nestory P. Gabagambi<sup>6</sup>, Alexandra M. Tyers<sup>7,9</sup>, Benjamin P. Ngatunga<sup>6</sup>, Milan Malinsky<sup>3,10</sup>, Richard Durbin<sup>2,3</sup>, George F. Turner<sup>7</sup>, Martin J. Genner<sup>4,11\*</sup> & Eric A. Miska<sup>1,2,3,11\*</sup>

<sup>1</sup>Wellcome/CRUK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK;

<sup>2</sup>Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK;

<sup>3</sup>Wellcome Sanger Institute, Hinxton CB10 1SA, UK;

<sup>4</sup>School of Biological Sciences, University of Bristol, Bristol BS8 1QU, UK;

<sup>5</sup>Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK;

<sup>6</sup>Tanzania Fisheries Research Institute (TAFIRI), Dar es Salaam, Tanzania;

<sup>7</sup>School of Natural Sciences, Bangor University, Bangor Gwynedd LL57 2UW, UK;

<sup>8</sup>Present address: School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, USA;

<sup>9</sup>Present address: Max Planck Institute for Biology of Ageing, Cologne 50931, Germany;

<sup>10</sup>Present address: Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland.

<sup>11</sup>Equal contribution.

\*Correspondence: M.J.G (m.genner@bristol.ac.uk), E.A.M. (eam29@cam.ac.uk), G.V (gv268@cam.ac.uk)

## ORCID

**GV:** 0000-0001-8942-2370  
**AGH:** 0000-0001-6436-2025  
**MES:** 0000-0003-3158-7935  
**AS:** 0000-0002-9228-0323  
**NPG:** 0000-0003-0260-2972

**MM:** 0000-0002-1462-6317  
**RD:** 0000-0002-9130-1006  
**GFT:** 0000-0003-0099-7261  
**MJG:** 0000-0003-1117-9168  
**EAM:** 0000-0002-4450-576X

## One sentence summary

Inheritance and plasticity of epigenetic divergence characterise early stages of speciation in an incipient cichlid species of an African crater lake.

## Abstract

Epigenetic variation can alter transcription and promote phenotypic divergence between populations facing different environmental challenges. Here we assess the epigenetic basis of diversification during the early stages of speciation. We focus on the extent and functional relevance of DNA methylome divergence between two *Astatotilapia calliptera* ecomorphs in crater Lake Masoko, southern Tanzania. We report extensive genome-wide methylome divergence between populations linked to key biological processes, including transcriptional activity of ecologically-relevant genes. These include genes involved in steroid metabolism, haemoglobin composition and erythropoiesis, consistent with divergent habitat occupancy of the ecomorphs. Using a common garden experiment, we found that global methylation profiles are rapidly remodelled across generations, but ecomorph-specific differences can be inherited. Collectively, our study suggests an epigenetic contribution to early stages of vertebrate speciation.

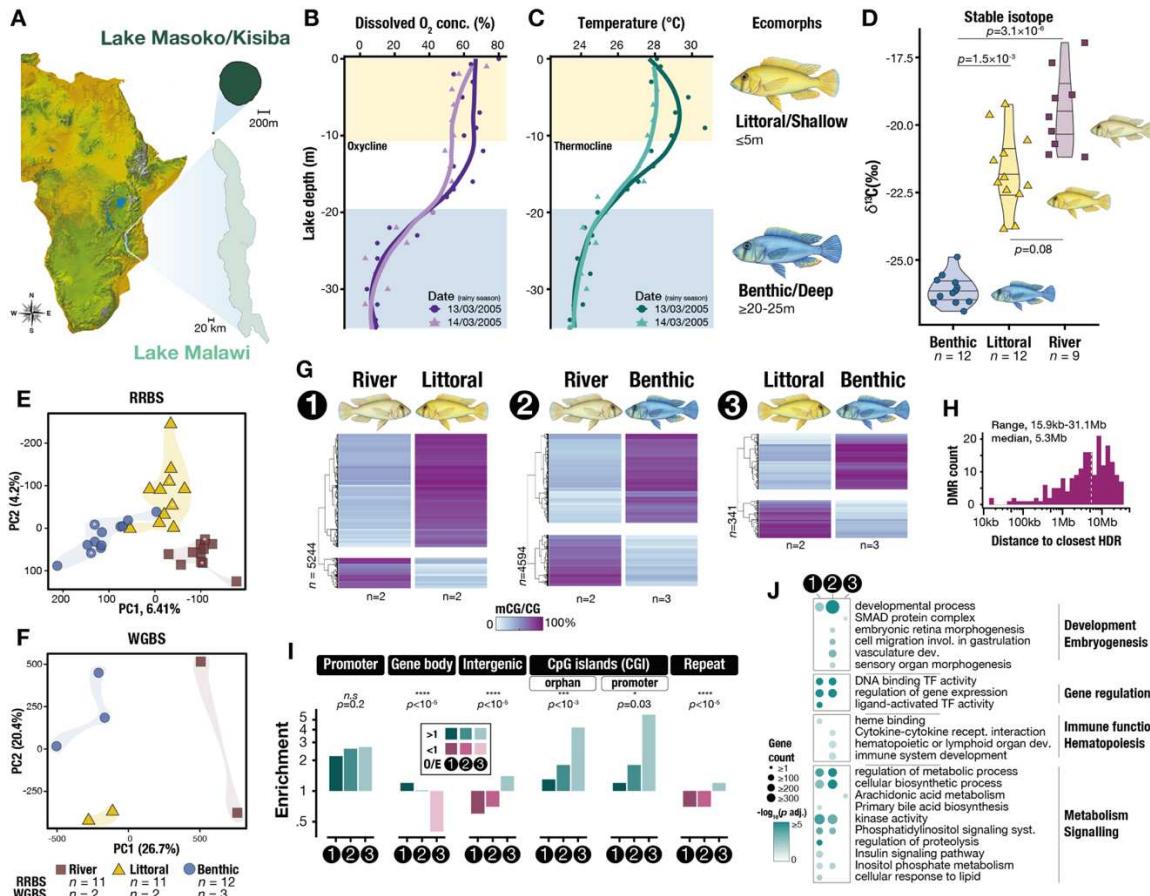
## Main text

The genomic basis of adaptive phenotypic diversification and speciation has been extensively studied but many questions remain (1–3). Recent studies in plants and animals have provided initial evidence for a contribution of heritable epigenetic divergence to functional phenotypic traits (4–9). However, whether epigenetic processes facilitate adaptive diversification, especially during the early stages of speciation, remains unknown.

To investigate the potential role of epigenetic processes during vertebrate speciation, we focus on the incipient *Astatotilapia* radiation in crater Lake Masoko, southern Tanzania, which is within the Lake Malawi catchment (Fig. 1A and Fig. S1A–C) (10). The two *Astatotilapia* ecomorphs present are characterised by different depth preferences, morphology, diet and male nuptial coloration (10, 11). Specifically, the littoral ecomorph (yellow males) occupies the well-oxygenated shallow waters ( $\leq 5$ m), while the benthic ecomorph (blue males) thrives in deeper (20–30m), almost anoxic and dimly lit habitats of the lake (Fig. 1B–C and Fig. S1C)(12). Previous genome-level sequencing has revealed overall low sequence divergence ( $F_{ST} = 0.038$ ) between the two ecomorph pairs (11), consistent with early stages of divergence. However, there was elevated genetic differentiation ( $F_{ST} > 0.3$ ) in 98 genomic regions, which were enriched for functional targets of divergent selection, notably in relation to vision, morphogenesis and hormone signalling (2, 11). Collectively, the two ecomorphs are monophyletic with respect to *Astatotilapia* in the neighbouring Mbaka river system (Fig. S1A), which they separated from ca. 10,000 years ago. More recently the colonisation and adaptation to the benthic habitat from littoral fish has taken place (1,000 years ago; 200–350 generations)(11). Against this background, the incipient *Astatotilapia* radiation of Lake Masoko provides a valuable system to investigate the role of epigenetic processes during the early stages of sympatric speciation. Here we combined reduced-representation bisulphite sequencing (RRBS), whole-genome bisulphite sequencing (WGBS) and whole transcriptome sequencing (RNAseq) to assess functional methylome divergence in the two *Astatotilapia* ecomorphs of Lake Masoko and their neighbouring ancestral riverine population.

### Population-specific methylome divergence during early stages of speciation

To examine genome-wide population-level methylome divergence, we combined two bisulphite sequencing approaches. First, we generated reduced-representation bisulphite sequencing (RRBS) data from liver tissue of 12 wild-caught adult males for each of the littoral and benthic *Astatotilapia* ecomorphs of Lake Masoko, and of 11 wild-caught adult males from the neighbouring Mbaka river. In addition, we produced whole-genome bisulphite sequencing (WGBS) data from liver tissue of at least two individuals from each population. On average,  $11.1 \pm 3.4$  million single-end 75bp-long reads and  $341 \pm 84.3$  million paired-end 150bp-long reads were generated per individual for the RRBS and WGBS dataset, respectively (Tables S1–S2; Methods). Reads yielded similar mapping rates to the SNP-corrected *Maylandia zebra* reference assembly across ecomorphs (GCF\_000238955.4; see Methods), consistent with low inter-population sequence divergence (Fig. S2A–B; see Materials and Methods)(11). Principal component analysis (PCA) of both RRBS and WGBS datasets revealed strong methylome segregation among the three populations (Fig. 1E–F).



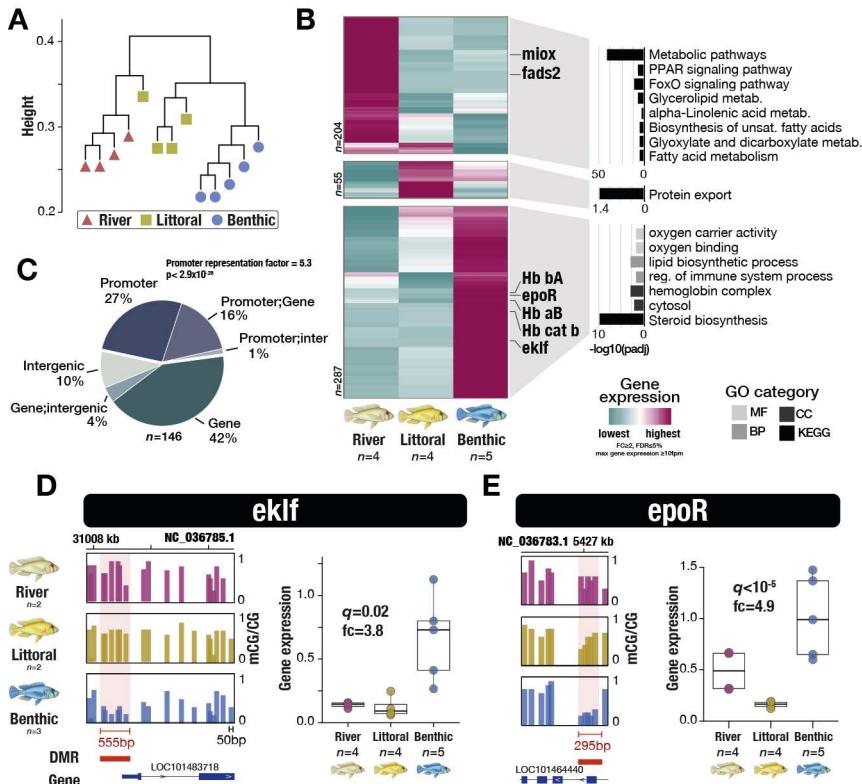
**Fig. 1. Whole-genome DNA methylation landscape of the *Astatotilapia* cichlid radiation in Lake Masoko. A.** Geographical map of Lake Masoko (also known as Lake Kisiba), Tanzania (satellite image: NASA). **B**. Dissolved oxygen ( $O_2$ ) concentration (%) by depth (meters, m) in Lake Masoko. **C**. Water temperature ( $^{\circ}C$ ) by depth. The oxycline (B) and thermocline (C) divide the habitats of the two *Astatotilapia calliptera* ecomorphs of Lake Masoko: the littoral (yellow) morph population thrives in shallow ( $\leq 5m$ ), well oxygenated waters, while the benthic (blue) morph population is found in deeper, colder and almost anoxic habitats of the lake. Data from Ref.(12), collected on two different days at same depths (LOESS smoothed curves). **D**. Violin plots of stable isotope ratios ( $\delta^{13}C_{VPDB}$ , ‰) by population indicate a more offshore zooplankton-based diet for the benthic fish.  $p$ -values for Dunn's multiple comparison tests are shown.  $n$  represents the number of fish per population. **E to F**. Principal component analysis (PC1-2) of liver methylome (mCG) variation using both the reduced representation bisulfite sequencing (RRBS)(E) and whole-genome bisulfite sequencing (WGBS)(F) datasets. PC1 and PC2 segregate the three populations apart in both datasets. Percentage of total variance given in parentheses. Asterisks in (E) indicate samples used for WGBS.  $n$  indicates the number of biological replicates for each population. **G**. Unbiased hierarchical clustering and heatmap of the average DNA methylation levels at all significantly differentially methylated regions found between the three pairwise comparisons (numbered 1-3) reveal population-specific methylome patterns. mCG/CG levels (%) averaged by population over each DMR ( $p \leq 0.05$ ). Total number of DMRs for each comparison shown vertically. **H**. Histogram of closest distance in bp (log scale; median, dotted line) between DMRs and HDRs (when on same chromosome). **I**. Enrichment plots (observed/expected ratio) for methylome variation (DMRs) in different genomic features for each pairwise comparison. Chi-square tests ( $p$  values shown) were performed for the three population distributions for each genomic feature. Repeat: transposon-only repeats. **J**. Gene ontology enrichment analysis for genes associated with methylome differences (either in promoter, gene body or 1-5kbp away from a given gene) for each pairwise comparison.

As functional methylation variation tends to occur over neighbouring CG dinucleotide sites, we can identify differentially methylated regions (DMRs;  $\geq 25\%$  methylation difference,  $\geq 4$  CpG,  $\geq 50$  bp-long,  $p < 0.05$ ; Materials and Methods) between each pair of populations using the WGBS dataset. In total, 5244 DMRs were found between the riverine and littoral fish, and 4595 between riverine and benthic fish (Fig. 1G and Fig. S3A). The majority of these DMRs (79% and 63.5% total DMRs, respectively) showed substantial gain in methylation (gain-DMRs;  $\geq 44\%$  methylation increase) in the littoral or benthic fish compared to ancestral riverine fish,

respectively, with average methylation levels at gain-DMRs of  $\geq 72\%$  mCG/CG (Fig. 1G and Fig. S3 B-C). Between the littoral and benthic fish, 341 DMRs were found, two-thirds of which showed an increase in methylation in the benthic fish population. In all pairwise comparisons, DMRs varied in length from 50 bp to 3 kbp (median, 250 bp) and were distributed across all chromosomes (Fig. S3 D-E). To investigate the relationship between underlying genetic polymorphism and methylome divergence, regions of elevated genetic differentiation (highly diverged region [ $F_{ST} \geq 0.3$ ], HDR; Ref. (11)) between the littoral and benthic populations were examined. These regions showed high methylome conservation and were not co-localised with DMRs. Notably, the distances between HDRs and the nearest DMRs ranged from 15.9 kb to 31.1 Mb (median, 5.3 Mb), suggesting that large regions of high genetic differentiation were in general not associated with epigenetic divergence between the two incipient Masoko ecomorphs (Figs. 1H and S3 G).

We next examined the genomic localisation of DMRs among populations, and found promoter regions to be highly enriched in all comparisons (~2.5-fold enrichment over chance; one sample t-test  $p < 0.0001$ ), consistent with their *cis*-regulatory functions (Fig. 1I) (13). In the recent littoral to benthic transition, the differences in methylomes were particularly overrepresented in CpG-dense regions (CpG-islands;  $\geq 4.5$ -fold enrichment;  $p < 0.0001$ ) located both within (pCGI) and outside promoters (orphan CGI/oCGI). oCGIs may represent distant *cis*-regulatory regions, such as ectopic promoters or enhancers (14), both of which are known targets of methyl-sensitive DNA-binding proteins (MDP) (13, 15). In contrast methylome variation within intergenic regions and transposable elements only showed slight enrichment in the comparison of the Masoko ecomorphs relative to comparisons involving the riverine population. Additionally, the methylome divergence in gene bodies, the function of which remains unclear in vertebrates but could be involved in alternative splicing (13), was generally very low, and even highly underrepresented (2.5-fold depletion,  $p < 0.0001$ ) in the comparison of Masoko ecomorphs. This suggests high methylome conservation between gene bodies of the most closely related populations.

We identified biological processes associated with methylome divergence by performing gene ontology enrichment analysis (Fig. 1J). We first noted that genes involved in transcription regulation were highly enriched in DMRs, and then identified three further biological processes for the genes enriched in methylome divergence: immune function and haematopoiesis; embryogenesis and development; and metabolism (Fig. 1J and Fig. S4). Notably, methylome divergence in developmental genes has been shown previously to account for close to half of all species-specific epigenetic differences between three species of the Lake Malawi cichlid radiation (7). Additionally, regions showing benthic-specific methylome patterns and located outside gene body were significantly enriched for specific transcription factor (TF) binding motifs (Fig. S3F), including the haematopoiesis-related SCL/Tal1 (stem cell leukaemia/T-cell acute lymphocytic leukaemia 1;  $p = 1 \times 10^{-145}$ ), forkhead hepatocyte nuclear factor 3-alpha (foxa1;  $p = 1 \times 10^{-9}$ ), the embryogenesis-related smad2 (mothers against decapentaplegic homolog 2;  $p = 1 \times 10^{-125}$ ) and the metabolic/circadian clock transcriptional activator bmal1 (aryl hydrocarbon receptor nuclear translocator-like protein 1;  $p = 1 \times 10^{-8}$ ), consistent with altered transcription factor activity. This suggests *cis*-regulatory functions for such population-specific DMRs in development, haematopoiesis and metabolism, associated with the adaptation to the benthic habitat.



**Fig. 2. Methyloome variation is associated with altered transcriptional activity.** **A.** Unbiased hierarchical dendrogram based on whole transcriptome variation (Euclidean distances plotted). **B.** Unbiased hierarchical clustering and heatmap of transcriptional activity (Z-score) for all significantly differentially expressed genes (DEGs) among the three wild populations, showing three different clusters of transcriptional activity (false discovery rate adjusted p-value, using Benjamini-Hochberg [FDR] <0.05, fold change  $\geq 2$  and high gene expression [top 90<sup>th</sup> percentile] in any one sample). Examples of differentially expressed genes are shown on the right-hand side of the graph. Right: gene ontology enrichment for DEGs of each of the three transcriptional activity clusters (only qval<0.01 shown). **C.** Pie chart representing the genomic localisation of DMRs associated with DEGs. Significant overlap between DMRs at promoter and transcriptional changes (representation factor= 5.3, p<10<sup>-28</sup>; exact hypergeometric probability). DEGs can be associated with multiple DMRs in different locations (promoter, intergenic and genic DMRs). **D-E.** The promoters of the genes erythroid transcription factor krüppel-like factor 1 (eklf) and the erythropoietin receptor (epoR), both involved in erythropoiesis and red blood cells differentiation, show hypomethylation levels in the livers of the benthic fish compared to the littoral populations. Genome browser view of methyloome profiles for each ecomorph are shown. Each bar represents the average mCG/CG levels in 50bp-long non-overlapping windows for each ecomorph group. DMRs (p<0.05) are highlighted in red and the length of each DMR is indicated in red. Right-hand side of D and E: boxplots of gene expression in liver of benthic, littoral and river fish for eklf and epoR are shown (qval, FDR; fc, fold change).

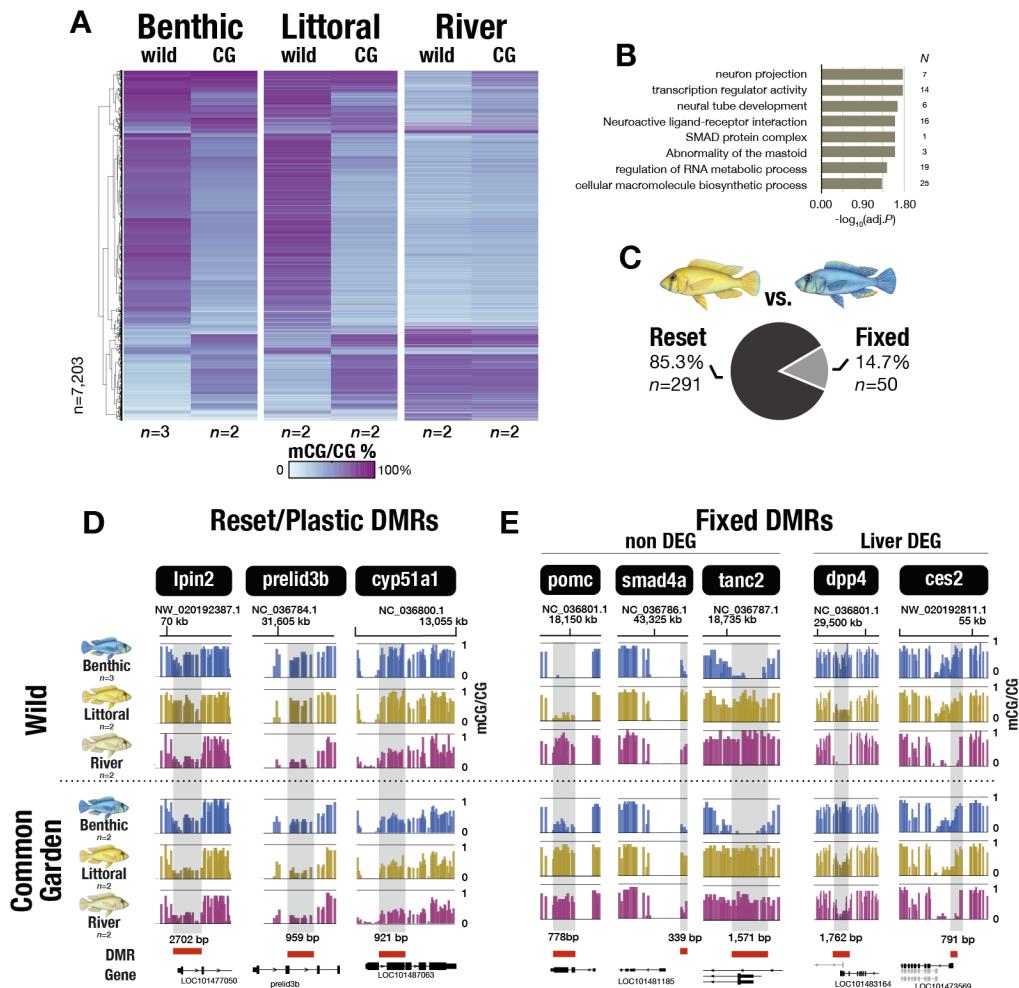
### Methyloome divergence associated with transcriptional changes in haematopoiesis and liver-specific metabolism

We investigated the link between population-specific methyloome divergence and transcriptional activity using total RNA sequencing data from liver tissues of 4-5 individuals of each group (mean  $\pm$  sd,  $32.9 \pm 3.8$  million paired-end 100/150bp-long reads per fish; Table S3). As with whole methyloome variation, we observe population-specific transcription patterns (Fig. 2A). We then performed differential gene expression analysis (false discovery rate [FDR] adjusted p-value using Benjamini-Hochberg < 1%, fold change  $\geq 1.5$  and high gene expression in  $\geq 1$  population [top 90<sup>th</sup> percentile]; see Materials and Methods) and found a total of 546 significantly differentially expressed genes (DEGs) between the three populations, including 125 genes that were differentially expressed between the Masoko ecomorphs (Fig. 2B). Close to 38% of all DEGs show reduced transcriptional activity in Masoko ecomorphs relative to the riverine population, and these genes were significantly enriched for functions

related to energy balance/homeostasis and steroid metabolism, including the PPAR (peroxisome proliferator-activated receptor) and foxO (forkhead box O) signalling pathways, consistent with metabolic adaptation to different diets (Fig.2B). Conversely, almost all the remaining DEGs (52%) show high transcriptional activity almost exclusively in the benthic fish and were primarily enriched for functions associated with haemoglobin complex/oxygen binding activities and iron homeostasis as well as fatty acid metabolism, in line with the occupation of a hypoxic environment, and possibly related to their zooplankton-rich diet (Fig.1 A, D). Critically, significant changes in transcriptional activity were strongly associated with methylome divergence at promoter regions (44% of all DEGs, overrepresentation factor=5.3,  $p<10^{-28}$ ; Fig. 2C).

Focusing on the functional categories identified above, we examined in detail several examples of transcriptional diversification associated with divergence in methylome landscapes. First, loss of methylation in two genes associated with active haematopoiesis and erythrocyte differentiation, the erythroid krüppel-like transcription factor (*eklf*) and the erythropoietin receptor (*epoR*), was associated with significant gain of transcriptional activity in the benthic fish compared to the littoral population (fold changes  $\geq 3.8$ ,  $FDR<0.02$ ; Fig. 2D-E). *eklf* plays an essential role in erythropoiesis, haem synthesis and in the modulation of globin genes expression (16), and its transcriptional activity has been linked to methylation changes previously in humans (17). Moreover, three haemoglobin subunit (Hb) genes, adjacent to each other within the MN (major) globin cluster on chromosome LG4 (NC\_036783.1) were significantly upregulated ( $\geq 4.5$ -fold) in the benthic fish specifically ( $FDR<0.02$ ; Figs. S5A). This includes the cathodic Hb beta (cHbB) which is known to have higher oxygen-binding affinity in anoxic environments and is expressed exclusively in benthic fish (18). Notably, a 2kbp-long DMR adjacent to this globin gene cluster shows benthic-specific hypermethylation methylome divergence compared to intermediate and unmethylated levels in littoral and river fishes respectively (Fig. S6). This suggests it bears *cis*-regulatory functions, possibly similar to the vertebrate Locus Control Regions (LCR), known to be bound by many essential erythroid transcription factors, such as *eklf*, *Scl/Tal1* and *Gata1* (19). Finally, the hematopoietic TF *Scl/Tal1*, a major actor in red blood cell differentiation (19), whose sequence-recognition binding motif is highly enriched in DMR sequences (Fig. S3F), is also expressed only in benthic fish ( $FDR<0.014$ ; Fig. S5B), suggesting altered TF activity associated with methylome differences in benthic fish. Collectively, these results suggest significant divergence in epigenetic and transcriptional landscapes affecting erythropoiesis and haemoglobin composition in the benthic fish, which may facilitate occupation of anoxic conditions within the benthic habitat.

As noted above, the Lake Masoko populations are mainly characterised by an overall reduced transcriptional activity in many genes related to steroid metabolic pathways and energy homeostasis compared to the riverine fish (Fig. 2B). Many of these genes show significant gain in methylation compared to the riverine population. The examples we examined include the fatty acid desaturase 2 (*fads2*), part of the PPAR signalling pathway, with functions in fatty acid metabolism and dietary metabolic adaptations (20), as well as the catabolic enzyme inositol oxygenase (*miox*). Both genes show hypermethylation associated with weaker transcriptional activity in Masoko ecomorphs compared to the ancestral-related riverine fish (Fig. S7A-B). The interplay between metabolism and epigenetic variation has been well established (8, 20, 21), and suggests that epigenetic divergence might facilitate any differences in dietary resource use patterns during colonisation of Lake Masoko habitats.



**Fig. 3. Common garden experiment results in global resetting of methylome profiles in wild benthic and littoral fish to resemble riverine methylome profiles, and inheritance of fixed methylome differences.** **A.** Heatmap of the average DNA methylation levels (mCG/CG, %) at all DMRs found between wild populations in Fig. 1G for all wild and common-garden (CG) fish. Methylome profiles reveal global epigenetic resetting in wild benthic and littoral fish, to resemble the ancestral-like, riverine methylome profiles. **B.** Gene ontology enrichment analysis showing enrichment for fixed DMRs in developmental genes. **C.** The proportion of reset (upon common garden experiment and within one generation) and population-fixed DMRs between littoral and benthic fish. **D-E.** Examples of DMRs reset upon common garden experiment (D) and associated with population-specific transcriptional differences, and of DMRs fixed between populations (E) in wild and common-garden fish. Some fixed DMRs were also associated with altered transcriptional activity between populations (DEG, differentially expressed genes; see Fig. S9F-H). Each bar represents the average mCG/CG levels in 50bp-long non-overlapping windows for each fish population. DMRs ( $p<0.05$ ) are highlighted in red and the length of each DMR is indicated in black.

### Plasticity and inheritance of population-specific methylome divergence

We examined the plasticity and inheritance of population-specific methylome divergence using a common garden experiment, whereby wild-caught *A. calliptera* specimens from Lake Masoko (both littoral and benthic populations), as well as from the adjacent river Mbaka, were bred and first-generation fish were reared under the same controlled laboratory conditions. Liver methylomes (WGBS) were then generated from two common-garden individuals for each population (Fig. S8A-B). Under a common rearing environment, and within one generation, the majority of DMRs (~86%) found in wild populations of Lake Masoko were reset to resemble mostly unchanged methylome levels of riverine-ancestral fish ('reset DMRs'), while the remaining ~15% were retained/fixed between populations ('fixed DMRs'; Fig. 3A-C and Figs. S8C and S9A), consistent with transgenerational retention of population-specific methylome patterns. Moreover, reset DMRs were on average

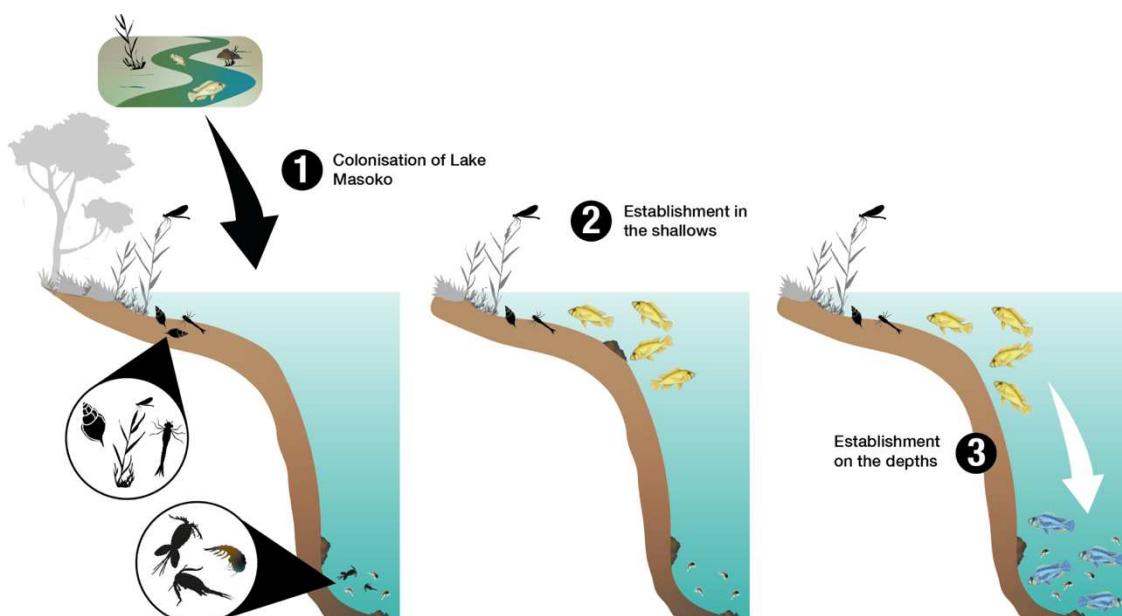
almost twice as long as fixed DMRs (median: 424 and 227 bp, respectively; Fig. S9I). We then performed gene ontology enrichment analysis and found that reset DMRs were significantly enriched in the promoter of genes with functions related to liver metabolism, immune function and with DNA-binding activity (Fig. S9B-C; see also Fig. S4). Moreover, 77.4% of all differentially expressed genes found among the livers of wild populations were associated with reset methylome patterns upon environmental perturbation (Fig. S9C). Such genes include for example phosphatidate phosphatase LPIN2 involved in fatty acid metabolism, PRELI domain containing protein 3B (preli3b) implicated in phospholipid transport, and lanosterol 14-alpha demethylase (cyp51a1) with functions in sterol biosynthesis (Fig. 3D), suggesting a tight link between environmental conditions and methylome divergence associated with altered metabolism. Furthermore, the erythropoietic *eklf* and *epoR*, which have benthic-specific methylome and transcriptome patterns in wild fish, all resemble river and littoral highly methylated profiles in the common garden experiment (Fig. S9D).

While the majority of population-specific methylome patterns in wild populations show high levels of plasticity, suggesting a tight link between environmental conditions and epigenetic variation, methylome patterns fixed between populations show an overall significant association with genes related to development and embryogenesis, in particular brain development (Fig. 3B). Fixed DMRs might represent population-specific tissue-independent methylome patterns, possibly reflecting distinct core developmental processes (7). Examples of genes with functions during brain development and neuron morphogenesis and showing fixed population-specific methylome patterns include the embryonic transcription regulator SMAD family member 4a (*smad4a*; hypomethylated in littorals only) and the scaffolding protein *tanc2* (hypomethylated in benthics) (Fig. 3E). Other pathways, related to biosynthetic processes among others, are also associated with fixed population-specific methylome patterns (Fig. 3B). This includes the gene pro-opiomelanocortin (*pomc*), a precursor polypeptide produced in the brain and involved in energy homeostasis and immune functions, showing retention of benthic-specific methylome patterns (Fig. 3E). Although three-quarters of all DEGs are associated with reset DMRs, we found some association between fixed population-specific methylome patterns and transcriptional divergence in liver-specific genes (22.6% of all DEGs), with functions enriched for catabolic and drug metabolism pathways (Fig. S10E). Two genes, in particular, coding for enzymes part of the insulin and fatty acid metabolic pathways respectively, *dpp4* (dipeptidyl peptidase 4) and *ces2* (liver carboxylase 2), show significant transcriptional downregulation in benthic livers and are associated with fixed benthic-specific hypermethylated levels at their promoters (Fig. 3E and Fig. S9G-H), suggesting retained epigenetic divergence associated with adaptation to different diets.

These results suggest that although methylome landscapes are highly environment-specific showing high plasticity and significantly associated with alteration of transcriptional activity of functional genes, some methylome divergence has become fixed and may be inherited in populations of Lake Masoko (fixed DMRs; Fig. 3D-E). While in mammals two waves in DNA methylation reprogramming occur early on, in zebrafish the paternal methylome is retained upon fertilisation, possibly allowing for transgenerational epigenetic inheritance (22) - however epigenetic reprogramming mechanisms in teleost fish may vary (23) and is currently unknown in cichlids.

## Conclusion

Our results provide direct evidence for functional and heritable methylome divergence associated with early stages of speciation in the radiation of *Astatotilapia* ecomorphs in Lake Masoko. Our analysis suggests that ancestral colonisation of the shallow lake habitat by a generalist riverine *Astatotilapia* population (~10,000 years ago) was followed by colonisation of benthic habitat (~1,000 years ago) characterised in part by both reversible and heritable methylome divergence in key functional genes (Fig. 4), including those related to haemoglobin synthesis, erythropoiesis, and sterol metabolism. We suggest that epigenetic processes provide the capacity for rapid occupation of new ecological niches prior to fixation of epigenetic and genomic variation. Our study therefore builds on evidence of epigenetic divergence seen among populations of fishes (24–26), intraspecific methylome remodelling in different rearing environments (27, 28), heritability of population-specific methylome divergence (25, 29, 30), and an epigenetic basis to diversification in vertebrate functional ecomorphological traits [eyes of cavefish *Astyanax mexicanus* (4)]. A key challenge now is to determine the mechanisms and rates of fixation of heritable epigenetic variation within populations (6, 8, 22), and how this associates with genomic fixation observed during the later stages of the speciation in cichlid fishes, and other model organisms used to study speciation and adaptive radiation (4–6, 8, 9).



**Fig. 4. Hypothesised three stages of epigenetically-associated *Astatotilapia* ecomorph speciation in Lake Masoko.** **1.** Ancestral colonisation of the shallow habitats of Lake Masoko by generalist riverine *Astatotilapia* population approximately 10,000 years ago (Malinsky et al., 2015). **2.** Occupancy of shallow, reedy highly oxygenated habitats by fish with a high level of depth philopatry. Phenotypic plasticity, partially linked to global methylome changes enables utilisation of littoral macroinvertebrate prey. **3.** Colonisation of deep, zooplankton-rich and low oxygen habitats by the shallow population ~1,000 years ago (Malinsky et al., 2015). Extreme methylome changes in the benthic population associated with diet (e.g., fatty acid metabolism) and environment (e.g., haemoglobin composition). Epialleles become reciprocally fixed in the two populations, plausibly leading migrants, and those of intermediate epi-genotypes to suffer a fitness disadvantage. Eventually, selection leads to differential fixation of genomic variation.

## REFERENCES

1. D. Brawand, C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J.-F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen, F. Di Palma, The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. **513**, 375–381 (2014).
2. H. Svardal, W. Salzburger, M. Malinsky, Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes. *Annu. Rev. Anim. Biosci.* **9**, 55–79 (2021).
3. W. Salzburger, Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**, 705–717 (2018).
4. A. V. Gore, K. A. Tomins, J. Iben, L. Ma, D. Castranova, A. E. Davis, A. Parkhurst, W. R. Jeffery, B. M. Weinstein, An epigenetic mechanism for cavefish eye degeneration. *Nat. Ecol. Evol.* **2**, 1155–1160 (2018).
5. T. Kawakatsu, S. C. Huang, F. Jupe, E. Sasaki, R. J. Schmitz, M. A. Urich, R. Castanon, J. R. Nery, C. Barragan, Y. He, H. Chen, M. Dubin, C.-R. Lee, C. Wang, F. Bemm, C. Becker, R. O'Neil, R. C. O'Malley, D. X. Quarless, N. J. Schork, D. Weigel, M. Nordborg, J. R. Ecker, C. Alonso-Blanco, J. Andrade, C. Becker, F. Bemm, J. Bergelson, K. Borgwardt, E. Chae, T. Dezwaan, W. Ding, J. R. Ecker, M. Expósito-Alonso, A. Farlow, J. Fitz, X. Gan, D. G. Grimm, A. Hancock, S. R. Henz, S. Holm, M. Horton, M. Jarsulic, R. A. Kerstetter, A. Korte, P. Korte, C. Lanz, C.-R. Lee, D. Meng, T. P. Michael, R. Mott, N. W. Mulyati, T. Nägele, M. Nagler, V. Nizhynska, M. Nordborg, P. Novikova, F. X. Picó, A. Platzer, F. A. Rabanal, A. Rodriguez, B. A. Rowan, P. A. Salomé, K. Schmid, R. J. Schmitz, Ü. Seren, F. G. Sperone, M. Sudkamp, H. Svardal, M. M. Tanzer, D. Todd, S. L. Volchenboum, C. Wang, G. Wang, X. Wang, W. Weckwerth, D. Weigel, X. Zhou, Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell*. **166**, 492–505 (2016).
6. E. A. Miska, A. C. Ferguson-Smith, Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. *Science*. **354**, 59–63 (2016).
7. G. Vernaz, M. Malinsky, H. Svardal, M. Du, A. M. Tyers, M. E. Santos, R. Durbin, M. J. Genner, G. F. Turner, E. A. Miska, Mapping Epigenetic Divergence in the Massive Radiation of Cichlid Fishes. *bioRxiv*, 1–25 (2020).
8. G. Cavalli, E. Heard, Advances in epigenetics link genetics to the environment and disease. *Nature*. **571**, 489–499 (2019).
9. P. Cubas, C. Vincent, E. Coen, An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*. **401**, 157–161 (1999).
10. G. F. Turner, B. P. Ngatunga, M. J. Genner, The Natural History of the Satellite Lakes of Lake Malawi. *EcoEvoRxiv* (2019), doi:10.32942/osf.io/sehdq.
11. M. Malinsky, R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, G. F. Turner, Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*. **350**, 1493–1498 (2015).
12. M. Delalande, thesis, Université Paris Sud - Paris XI (2008).
13. D. Schübeler, Function and information content of DNA methylation. *Nature*. **517**, 321–326 (2015).
14. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–22 (2011).
15. Y. Yin, E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schübeler, C. Vinson, J. Taipale, Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. **356**, eaaj2239 (2017).
16. M. R. Tallack, T. Whittington, W. S. Yuen, E. N. Wainwright, J. R. Keys, B. B. Gardiner, E. Nourbakhsh, N. Cloonan, S. M. Grimmond, T. L. Bailey, A. C. Perkins, A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res.* **20**, 1052–1063 (2010).
17. Y. Li, D. Liu, Z. Li, X. Zhang, Y. Ye, Q. Liu, J. Shen, Z. Chen, H. Huang, Y. Liang, X. Han, J. Liu, X. An, N. Mohandas, X. Xu, Role of tissue-specific promoter DNA methylation in regulating the human EKLF gene. *Blood Cells, Mol. Dis.* **71**, 16–22 (2018).
18. J. F. Storz, Gene Duplication and Evolutionary Innovations in Hemoglobin-Oxygen Transport. *Physiology*. **31**, 223–232 (2016).
19. P. P. Levings, J. Bungert, The human  $\beta$ -globin locus control region. A center of attraction. *Eur. J. Biochem.* **269**, 1589–1599 (2002).
20. H. Xu, X. Dong, Q. Ai, K. Mai, W. Xu, Y. Zhang, R. Zuo, Regulation of Tissue LC-PUFA Contents,  $\Delta 6$  Fatty Acyl Desaturase (FADS2) Gene Expression and the Methylation of the Putative FADS2 Gene Promoter by Different Dietary Fatty Acid Profiles in Japanese Seabass (*Lateolabrax japonicus*). *PLoS One*. **9**, e87726 (2014).
21. V. M. Sales, A. C. Ferguson-Smith, M. E. Patti, Epigenetic Mechanisms of Transmission of Metabolic Disease across Generations. *Cell Metab.* **25**, 559–571 (2017).
22. K. Skvortsova, N. Iovino, O. Bogdanović, Functions and mechanisms of epigenetic inheritance in animals. *Nat. Rev. Mol. Cell Biol.* **19**, 774–790 (2018).

23. X. Wang, R. K. Bhandari, DNA methylation dynamics during epigenetic reprogramming of medaka embryo. *Epigenetics*. **14**, 611–622 (2019).
24. T. A. Smith, M. D. Martin, M. Nguyen, T. C. Mendelson, Epigenetic divergence as a potential first step in darter speciation. *Mol. Ecol.* **25**, 1883–1894 (2016).
25. J. Hu, S. J. S. Wuitchik, T. N. Barry, H. A. Jamniczky, S. M. Rogers, R. D. H. Barrett, Heritability of DNA methylation in threespine stickleback ( *Gasterosteus aculeatus* ). *Genetics*. **217**, 1–45 (2021).
26. T. Ryu, H. D. Veilleux, J. M. Donelson, P. L. Munday, T. Ravasi, The epigenetic landscape of transgenerational acclimation to ocean warming. *Nat. Clim. Chang.* **8**, 504–509 (2018).
27. D. Anastasiadi, N. Díaz, F. Piferrer, Small ocean temperature increases elicit stage-dependent changes in DNA methylation and gene expression in a fish, the European sea bass. *Sci. Rep.* **7**, 12401 (2017).
28. J. Le Luyer, M. Laporte, T. D. Beacham, K. H. Kaukinen, R. E. Withler, J. S. Leong, E. B. Rondeau, B. F. Koop, L. Bernatchez, Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proc. Natl. Acad. Sci.* **114**, 12964–12969 (2017).
29. M. J. Heckwolf, B. S. Meyer, R. Häslер, M. P. Höppner, C. Eizaguirre, T. B. H. Reusch, Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Sci. Adv.* **6**, eaaz1138 (2020).
30. J. L. Kelley, M. Tobler, D. Beck, I. Sadler-Riggleman, C. R. Quackenbush, L. Arias Rodriguez, M. K. Skinner, Epigenetic inheritance of DNA methylation changes in fish living in hydrogen sulfide-rich springs. *Proc. Natl. Acad. Sci.* **118**, e2014929118 (2021).

**ACKNOWLEDGMENTS:** We thank the staff of the Tanzania Fisheries Research Institute for their assistance and support. The Tanzania Commission for Science and Technology provided research permits, while the Tanzania Ministry of Livestock and Fisheries provided export permits. We thank Msafiri Ndawala (TAFIRI) and George Kazumbe for their help with sample collection, and Julie Johnson for her fish paintings. We thank Navin B. Ramakrishna for critical comments on the manuscript. We thank the staff at the Gurdon Institute and the sequencing facilities at CRUK Cambridge Institute, Wellcome Gurdon and Wellcome Sanger Institutes, and Genomics facility at the University of Bristol for their expertise and support.

**FUNDING:** This work was supported by the following grants to EAM: Wellcome Trust Senior Investigator award (104640/Z/14/Z and 219475/Z/19/Z) and CRUK award (C13474/A27826); to RD: Wellcome award (WT207492); to GFT: Leverhulme Trust Award RPG-2014-214; to MJG, GFT and BPN: the Leverhulme Trust - Royal Society Africa Awards (AA100023 and AA130107); to MJG, GFT: NERC award (NE/S001794/1); to MJG: Leverhulme Trust award (RF-2014-686); to AGH: Marie Skłodowska-Curie Individual Fellowship (GA-659791). GV acknowledges Wolfson College, University of Cambridge and the Genetics Society, London for financial support. The authors also acknowledge core funding to the Gurdon Institute from Wellcome (092096/Z/10/Z, 203144/Z/16/Z) and CRUK (C6946/A24843). For Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

**AUTHOR CONTRIBUTION:** Sample collection: GV, AGH, AS, BPN, NPG, MC, MES, AMT. Common Garden experiment: GFT. RNA extraction: GV, BF. RRBS libraries and analysis: AGH, GV. WGBS libraries: GV. HDR annotation, MM; WGBS, RNAseq, RRBS analysis: GV. Supervision: MJG, EAM, RD, GV, MES. Conceptualisation: MJG, GFT, EAM, GV. Writing: GV, MJG, EAM, with contribution from all the authors.

**COMPETING INTEREST:** None declared.

**DATA AND MATERIALS AVAILABILITY:** WGBS, RRBS and RNA sequencing raw data have been deposited in the Gene Expression Omnibus (GEO) database and will be made public upon publication.

**SUPPLEMENTARY MATERIALS:**

Material and Methods

Supplementary Figures S1-S9

Supplementary Tables S1-S3