Research article

Title: Genetic and epigenetic interplay allows rapid transgenerational adaption to metal pollution in zebrafish

List of authors:
Fabien Pierron[a*], Débora Heroin[a], Guillemine Daffe[b], Flore Daramy[a], Aurélien Barré[c], Olivier Bouchez[d], Alicia Romero-Ramirez[a], Patrice Gonzalez[a] and Macha Nikolski[c,e]

Authors' affiliation adresses:
[a]Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France
[b]Univ. Bordeaux, CNRS, INRAE, La Rochelle Univ., UMS 2567 POREA, F-33615 Pessac, France
[c]Univ. Bordeaux, Bordeaux Bioinformatics Center, Bordeaux, France
[d]INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, 31326, France
[e]Univ. Bordeaux, CNRS, IBGC, UMR 5095, Bordeaux, 33077, France

***Corresponding author:**
Fabien Pierron
UMR 5805 EPOC - OASU
Station Marine d'Arcachon
Université de Bordeaux
Place du Docteur Bertrand Peyneau
33120 ARCACHON
France
Phone : +335 56 22 39 33
fabien.pierron@u-bordeaux.fr
Orcid
Fabien Pierron 0000-0001-6176-3050

Abstract

Despite still being a matter of debate, there is growing evidence that pollutant-induced epigenetic changes can be propagated across generations. Whereas such modifications could have long lasting effects on organisms and even on population, environmentally relevant data from long-term exposure combined with follow-up through multiple generations remain scarce for non-mammalian species. We performed a transgenerational experiment comprising 4 successive generations of zebrafish. Only fish from the first generation were exposed to an environmentally realistic concentration of cadmium (Cd). Using a whole methylome analysis, we first identified the DNA regions that were differentially methylated in response to Cd exposure and common to fish of the first two generations. Among them, we then focused our investigations on the exon 3 of the *cep19* gene. We indeed recorded transgenerational growth disorders in Cd exposed fish and a mutation in this exon is known to cause morbid obesity in mammals. Its methylation level was thus determined in zebrafish from all the 4 generations by means of a targeted and base resolution method. We observed a transgenerational inheritance of Cd-induced DNA methylation changes up to the 4th generation. However, these changes were closely associated with genetic variations, mainly a single nucleotide polymorphism. This SNP was itself at the origin of the creation or deletion of a methylation site and deeply impacted the methylation level of neighboring methylation sites. Cd-induced epi-genetic changes were associated with different mRNA transcripts and an improved condition of Cd fish. Our results emphasize a tight relationship between genetic and epigenetic mechanisms and suggest that their interplay and pre-existing diversity can allow rapid adaptation to anthropogenic environmental changes.

Key words: Transgenerational growth disorders, DNA methylation, SNP, pollution, rapid adaptation

1. Introduction

The fact that stress-induced changes in epigenetic marks can not only have long-lasting effects on organisms but also on their offspring has prompted renewed and growing interest of biologists for epigenetics in the last decades [1-2]. Especially, the fact that these changes can be inherited across several generations raises the possibility that epigenetic mutations could not only contribute to heritable phenotypic traits and transgenerational diseases but even to evolution by acting as precursors of genetic mutations [3-4]. However, even if this mode of inheritance is well recognized in plants and despite the fact that examples of epigenetic inheritance in different taxa have multiplied in the last years, there is still a controversy as to whether transgenerational inheritance is widespread, particularly in vertebrates [5-7]. By controlling the activity of genes, epigenetic mechanisms play a crucial role in cell differentiation during development [1,3,8]. Cells differentiate, acquire and maintain identity by expressing a particular transcriptome [9]. Thus, the epigenome must be re-programmed at least to ensure totipotency after fertilization. In vertebrates, particularly in mammals, the epigenome does not persist throughout life, but rather undergoes precise and coordinated changes at precise stages of development. The two rounds of DNA methylation erasure that occur in early embryo and germline leave little chance to the inheritance of stress-induced changes in epigenetic marks [6,10]. Furthermore, another critical argument against epigenetic inheritance comes from the fact that genetic variations are rarely or even never taken into account in transgenerational epigenetic studies; i.e. some examples of transgenerational epigenetic inheritance could be explained by changes in the DNA sequence itself. Finally, true epigenetic inheritance can only be determined after several generations and reported examples of epigenetic inheritance in vertebrates often concern inter- rather than true transgenerational effects [2,5-6].

Most of the current knowledge relating to pollutant-induced epigenetic modifications comes from rodent models and experimental investigations during which exposure to

pollutants was carried out during embryogenesis. Data from long-term exposure, which are environmentally more relevant, for non-mammalian species and for at least three consecutive generations remain scarce, although such modifications could have long lasting effects on organisms and even on the evolution trajectories of populations [3,5,11-15]. In view of these considerations, recent studies showed that non-mammalian vertebrate species such as fish can be more prone to transgenerational inheritance of stress-induced DNA methylation marks [16-17]. Although DNA methylation is present in different phyla, it is in vertebrates and plants that this epigenetic mechanism has the most prominent role in transcriptional regulation and the greatest potential to contribute to transgenerational inheritance [1].

In this context, the aims of the present study were to test whether (i) pollutant exposure can induce DNA methylation changes in zebrafish and (ii) such pollutant-induced changes can be transmitted to the unexposed offspring at least until the fourth generation. To test these hypotheses, one generation of zebrafish (F0) was exposed throughout its life, from fertilization until sexual maturity, to dissolved cadmium (Cd) at an environmentally relevant concentration of 1 $\mu g.L^{-1}$ (see Pierron et al., 2021 [15]). Cd is a widespread nonessential metal with endocrine disrupting effects [15,18]. Zebrafish were life-long exposed to Cd to encompass the remodeling events of DNA methylation that occur in germline during the embryo-larval stages but also later in adult gonads at the onset of "puberty" [11,16-17,19]. We first used a high throughput pyrosequencing approach to identify without *a priori* the regions of DNA that were differentially methylated (called Differentially Methylated Regions, DMRs) in response to Cd exposure and common to fish of the F0 and F1 generation. Considering the Cd-induced transgenerational effects on zebrafish growth and condition, we then focused our investigations on a DMR identified in the exon 3 of the gene encoding for the ciliary protein CEP19 [20-21]. Its methylation level was determined in zebrafish from all the 4 generations by means of a targeted and base resolution method.

2. Material and methods

2.1 General rearing and breeding program

The experiment is presented in Fig. 1 and described in details in text S1. At the beginning of the experiment, 40 randomly mating pairs (one male and one female) of wild-type (domesticated) zebrafish were placed for one night in small breeding tanks, 20 tanks were filled with clean water (Control, C) and 20 tanks were filled with Cd-contaminated water (1 µg.L^{-1}). Their offspring (F0 generation) were respectively reared in clean or Cd-contaminated water throughout their life. For DNA methylation and transcription analyses, C and Cd fish were sampled at 5 dpf (larvae) and at 124 dpf (adult mature fish). Whole larvae and tissue samples of adults were fixed in RNAlater solution before being stored at -20°C. At 170 dpf, 20 pairs per condition were constituted to produce the next generation (F1). Mating was carried out in Cd-free water for both groups (C or Cd). Pairs were not randomly selected. For each condition, fish were reared in four different tanks. As all individuals from one clutch (i.e. from one pair) were raised in the same large tank, one male from one tank was mated with a female from another tank to limit inbreeding. With the exception of the fish that were used as genitors, all remaining individuals after mating were measured (total and standard (from snout to the base of the caudal fin, length) and weighted to determine the Fulton condition factor [K = (total weight (g)) / (total length (cm))^3) x 10^5] and investigate the length-weight relationship. After mating, genitors were conserved in one large tank under Cd-free condition where C and Cd fish were separated by a plastic net. Throughout the experiment, fish were fed three times daily. The food quantity delivered per aquarium was adjusted to fish density in order to not induce bias in growth. The same protocol as previously

described, was used for mating and sampling for the F1, F2 and F3 generations. However, for both groups of animals (C and Cd), spawning and rearing were carried out in Cd-free water. Thus, only fish of the F0 generation were directly exposed to dissolved Cd. The total duration of the experiment was 643 days.

This work was conducted in accordance with the EU Directive 2010/63/EU for animal experiments and received approval for research ethics from the Aquitaine fish-birds ethic committee (APAFIS#7535-2016111009351504).

2.2 RNA/DNA extraction

For each generation, a total of 6 biological replicates were used per condition. We used 6 larvae (5 dpf) from 6 different mating pairs and 6 adult males (124 dpf) and 6 adult females (124 dpf) from 3 different tanks in order to analyze individuals from different families. Only larval samples for the MeDip-Seq analysis were differently treated. Indeed, the amount of DNA required for MeDip-Seq compelled us to pool the DNA from 2 larvae (n = 6 pools per condition for a total of 12 larvae per condition). For all other samples and all other analyses (BS-Seq and RT-qPCR), no pool was used. Total RNA and DNA were purified from whole larvae or liver, brain and gonads of adults (males and females) using the AllPrep DNA/RNA kit (Qiagen) as described in Pierron *et al.* (2019) [22].

2.3 Identification of intergenerational Cd-induced DMRs

Libraries preparation and sequencing.

The MeDip-Seq (Methylated DNA immunoprecipitation coupled with high throughput pyrosequencing) approach was carried out on DNA samples from larvae and adult zebrafish of the F0 and F1 generation, for a total of 168 samples. Libraries were prepared according to

the Biooscientific's protocol using the Biooscientific NEXTflex™ Methyl-seq Library Prep Kit for Illumina Sequencing. Briefly, DNA was fragmented by sonication on a covaris M220 (400-500 bp), size selection was then performed using AMPure XP beads and adaptors were ligated. Libraries were then immunoprecipitated using the MagMeDIP Kit (Diagenode) according to manufacturer's instructions. Finally, 12 cycles of PCR were performed. Library quality was assessed using an Advanced Analytical Fragment Analyser and libraries were quantified by qPCR using the Kapa Library Quantification Kit. Following quantification, libraries were randomly pooled in equal amounts and sequenced on two lanes of an Illumina NovaSeq 6000 using a paired-end read length of 2x150 bp using the Illumina NovaSeq Reagent Kits. Sequencing was performed at the GeT-PlaGe core facility, INRAE Toulouse.

Data preprocessing and analysis. Sequencing files were quality checked using fastp (version 0.19.5). Filtered sequences were then mapped to the *Danio rerio* reference genome GRCz11 using the mem algorithm of the BWA software (version v 0.7.17) with default settings. The output files (*.sam) were then combined and converted to bam format (*.bam) by Samtools (version 1.4). Output files were grouped in 28 sets corresponding to the developmental stage, sex, generation (F0 or F1), organs studied and exposure. Analysis was carried out using the MEDIPS package (version 1.46.0) in R (version 3.5.1), following the MEDIPS R script. CpG density dependent normalization was performed on the mapping files for each set. No set was considered as a reference. The parameters specified in this analysis were set to: 'extend' set to 300, 'shift' set to 0, 'window size' set to 100, 'uniq' set to 1. Differentially methylated windows of 100 bp were identified between sets using the 'edgeR' plug-in function. For all comparisons, genomic windows that were covered by a minimal count threshold of 10 reads were kept for further statistical analysis. Statistically significant results were selected using criteria for the adjusted P-value ≤ 0.01. Contiguous significant windows were merged into one differentially methylated region (DMR). Based on the

biological questions, the DMRs were compared between the relevant sets. Since the final merging step can produce different DMRs between each pair of set, to enable pairwise comparisons, DMRs have been split in shared regions of size 300 or more. All DMRs having a positive or a negative fold change, were retained.

2.4 Analysis of single base methylation levels and single nucleotide polymorphism in cep19

Genomic sequences were obtained from the Genome Reference Consortium Zebrafish 11 (GRCz11). Specific primers used for PCR amplification and bisulfite sequencing were designed in the promoter-exon1 (ex1) region and exon 3 (ex3) of the *cep19* gene by means of the Pyromark assay design software (Table S1).

Quantification of single cytosine percent methylation at specific sites was performed as previously described in Pierron et al. (2019) by bisulfite-pyrosequencing using the PyroMark Q48 Autoprep instrument (Qiagen). SNP detection was carried out during the run of bisulfite pyrosequencing by modifying the sequence analyzed (AQ mode, Fig. S1). In order to assess the conversion efficiency of the bisulfite treatment, the average methylation level of one CpA site was measured in ex3. The methylation level of this control site was < 7 %, indicating very good conversion efficiency. The methylation level of 10 CpG sites were quantified in the ex3. The mean methylation level of these 10 CpG sites was used for statistical analyzes. Examples of pyrograms are provided in Fig. S1. To confirm our results, the same protocol, with different primers (Table S1), was used to sequence the ex3 by using native, untreated, genomic DNA as template (Fig. S2).

2.5 Cloning and sequencing of genetic/transcript variants

In order to identify potential transcript variants, several primers overlapping different exons were designed using the primer3plus software (Table S1, [23]). Additional variants of the *cep19* mRNA were identified using 3'RACE. First-strand cDNA was synthesized using the GoScript™ Reverse Transcription System (Promega) and an anchored oligo-dT (480 nM final, Table S1). A touch-down strategy was used as follows: 94 °C for 15 minutes; 12 cycles of 94 °C for 30s, 64.8-60.4 for 30s and 72 °C for 3 minutes; 35 cycles of 94 °C for 30s, 60°C for 30s and 72°C for 3 minutes and final extension at 72 °C for 10 minutes using specific forward primers and a reverse primer specific to the anchored oligo-dT primer (Table S1). Products were separated and visualized by gel electrophoresis or by means of a Qiaxcell instrument using the DNA high resolution kit (Qiagen). Amplified products of interest were cloned into pGEM-T Easy vector according to the manufacturer's instructions (Promega) and were sequenced (Sanger, GATC Biotech).

2.6 Quantitative analysis of transcription levels

RT-qPCR analyses were carried out as previously described in Pierron et al. (2019) [22] with few modifications. First-strand cDNA was synthesized from 250 ng of total RNA from individual whole larvae and was diluted 20-fold. Relative quantification of gene transcription was achieved by concurrent amplification of the *eef1a1a* and *rpl13a* endogenous controls [24]. As the same quantity of total RNA was used for each sample, the output cycle of the endogenous controls was statistically examined. No significant difference was observed among conditions and generations, demonstrating the relevance of the two genes as reference in our conditions. As the Ct value of the *eef1a1a* gene showed the lowest variability among samples, this gene was finally used as reference. Amplification efficiencies for all primer sets were calculated and used to determine transcription levels of targets. Primers used are available in Table S1.

2.7 Statistical analyses

Comparisons among fish groups were performed by two-ways analysis of variance (ANOVA), after checking assumptions of normality (Kolmogorov-Smirnov) and homoscedasticity of the error terms (Levene). If significant effects were detected, the LSD test was used to determine whether means between pairs of samples were significantly different from one another. When the assumptions were not met as deduced from ad-hoc tests, we used box-cox data transformations or the nonparametric Kruskal−Wallis test. Using Kruskal−Wallis, the post-hoc Conover-Iman test was used. Comparisons of proportions were performed using the Chi squared test and the Monte Carlo method using 5000 permutations followed by the Marascuilo procedure to compare proportions among the different conditions. Comparisons of allelic frequencies between groups were carried out by means of the homogeneity chi-square test. Comparisons of slopes were realized using a Student's t-test followed by the Bonferroni correction. Computations were performed using STATISTICA version 6.1 software (StatSoft) and XLSTAT (Addinsoft version 2020.1.1). Numerical results are provided as mean ± SE.

3. Results and discussion

3.1 Direct and transgenerational effect of Cd on fish growth

Cd exposure triggered an increase in fish growth in F0 generation. At 170 days post-fertilization (dpf), fish directly exposed to Cd were significantly heavier (Fig. 2A) and longer (Fig. 2B) than control specimens. However, they presented a Fulton condition factor that was significantly lower in comparison to controls (Fig. 2C). Fulton is an index frequently used in

ecology to assess the relative plumpness (i.e. energy reserves) of fish and corresponds to a weight to length ratio comparable to the body mass index in humans. After mating (170 dpf), C and Cd genitors were placed in one large tank under Cd-free condition where fish from the two conditions were separated by a plastic net. In the case of Cd fish, the weight gain strongly increased after 170 dpf at a rate that was almost exponential (Fig. 2D). However, as it can be observed for controls, the growth rate of zebrafish normally slows down with age [25]. At 400 dpf, Cd fish were almost obese (Fig. 2E) and were significantly 1.7-fold heavier than controls. At the opposite, in the subsequent generations (F1-F3), fish ancestrally exposed to Cd presented a body weight and length that were significantly lower than controls (Fig. 2A,B). However, from the generation F2, they presented a Fulton condition factor that was equivalent to controls (Fig. 2C). This highlights that Cd triggered transgenerational growth disorders. Whereas Cd exposure at the F0 generation initially boosted the growth of fish, the growth of their descendants (up to the 4th generation) was significantly lower in comparison to controls. However, their condition (i.e. Fulton) reached that of the controls starting from the 3rd generation.

3.2 Cd exposure triggered intergenerational changes in the methylation level of cep19

In order to identify without *a priori* the DNA regions that could be differentially methylated in response to Cd exposure and to maximize the number of samples analyzed, we first used a MeDip-Seq approach on both larvae and adult zebrafish of the F0 and F1 generation. Only few DMRs were common to the two generations and only 13 Cd-induced DMRs were common between larvae of the two generations (Table S2). Eleven of them were localized in non-coding regions, including intergenic and intronic regions. Only one DMR was found purely exonic. This DMR, that was localized on the exon 3 (ex3) of the gene encoding for the ciliary protein CEP19, retained our attention. Indeed, a homozygous

nonsense mutation in this exon causes morbid obesity in humans and mice [20-21]. As Cd exposure was associated with transgenerational growth disorders that were initially marked by an increased growth leading to signs of obesity in fish of the F0 generation, we then decided to investigate the methylation level of this gene throughout the 4 generations by means of a targeted and base resolution method, i.e. bisulfite pyrosequencing.

3.3 Transgenerational changes in the methylation level of the exon 3 of cep19 was associated with a single nucleotide polymorphism

Significant differences in the methylation level of the ex3 of the *cep19* gene were observed between C and Cd larvae of the F1 and F3 generation (Fig. 3A). Methylation levels were significantly higher in Cd larvae in comparison to C larvae. Generally, whatever the generation, higher methylation levels were observed in Cd fish in comparison to controls. However, high variations were observed among individuals, and methylation levels ranged from 4.3 to 90.4 %. No effect was observed on this region. We also identified a SNP (position 2941538 in NC_007133.7, GRCz11) during sequencing runs, where the A base, that is referenced in the RefSeq genome (GRCz11), was replaced by a G base. Interestingly, the presence of the alternative G base is at the origin of an additional site of methylation as this SNP is preceded by a C (CA→CG), methylated cytosines being primarily found at CpG dinucleotides [26]. Presence of this SNP deeply impacted the methylation level of neighboring CpG sites at both the larval and adult stages. Combining data from all analyzed individuals (larvae and adults of the four generations), the mean methylation level of the ex3 reached 7.08 ± 0.41 % (mean ± SE, n = 84) in fish that were homozygous for the A base (A/A genotype), 41.68 ± 0.64 % (mean ± SE, n = 47) in heterozygous A/G individuals and 83.48 ± 1.86 % (mean ± SE, n = 13) in homozygous G/G animals (Fig. 3B; raw data for the 10 CpG sites analyzed are available in Fig. S3). Considering that the detection limit of the instrument

is of 5%, methylation level in A/A individuals can be considered as negligible. Thus, our results show that CpG sites of the A allele are completely demethylated whereas CpG sites of the G allele are hypermethylated (Fig. 3C). In addition, as it can be observed in Fig. 3B, if highly significant changes in DNA methylation levels were observed among fish genotypes, no significant difference was observed between C and Cd fish of each genotype. These results highlighted a very strong relationship between the observed SNP and the methylation level of the exon. As promoter DNA methylation is an important regulatory component of vertebrate gene expression, we also measured the methylation level of the promoter region of *cep19*. No significant variation was observed in this region (Fig. S4). It is worth noting that the base change in the ex3 does not induce change in the protein sequence as the two different codons encode for the same amino acid. We then used the PROMO software (version 3.02, dissimilarity rate 5%; [27-28]) in order to identify potential binding sites for transcription factors (TFs) in the coding region of the ex3. Indeed, despite the widespread assumption that such binding in coding regions is non-functional, there are increasing evidences that coding exons are commonly bound by enhancer-related and histone modifying TFs and that many genes, mainly developmental genes, contain enhancers in coding regions that drive tissue-specific gene expression [29]. In addition, many studies have reported changes in gene expression in relation to synonymous SNPs in coding regions [30]. Finally, the chromatin structure of the genome is defined by a tight and dynamic interplay between histone modifications and DNA methylation [31]. In our case, the A→G transition triggered the loss of the binding site for the glucocorticoid receptor (GR also known as NR3C1). Interestingly, several papers have reported that binding of GR leads to the demethylation of its target genes [32-34]. Emerging evidences also suggest that TFs could hinder the methylation of their bound loci during global re-methylation that arises during embryogenesis [35]. In our case, the ex3 was effectively demethylated in the presence of the GR binding site (A allele). GR

mediates the diverse physiologic effects of glucocorticoids (GCs). GCs, of which cortisol is the best known, are steroid hormones that are important for maintenance of basal and stress-related homeostasis with significant effects on growth and weight maintenance [36-37]. Indeed, certain cases of obesity in humans have clear features of hypercortisolism [38-39]. Moreover, Cd exposure during pregnancy in mammals is well known to reduce birth weight, an effect mediated by an increase in maternal and fetal GCs [40]. In zebrafish, Wilson et al. (2016) [36] investigated whether embryo-larval GC exposure (0-5 dpf) affects adult zebrafish phenotypes. They experimentally induced an excess (by pharmacological treatment) or inversely a reduced (using an antisense GR morpholino) GC activity in zebrafish from 0 to 5 dpf. These treatments triggered respectively an increase and a decrease in the Fulton condition factor of fish at the adult stage (84 dpf).

This tight association between genetic and epigenetic changes can explain the high variations observed in DNA methylation levels (Fig. 3A). Each group was indeed composed of fish of different genotypes, i.e. of fish presenting differing levels of methylation. We then investigated the proportion of fish of each genotype per condition (Fig. 3B; i.e. for each condition, the number of fish of each genotype divided per the total number of fish analyzed x 100). The great majority of C fish (72.2 %) presented the A/A genotype. In contrast, Cd fish were equally comprised by fish of A/A and A/G genotypes (45.1 and 43.7%, respectively). Moreover, the proportion of fish presenting the A/A genotype was significantly lower in the Cd in comparison to the C condition. The proportion of homozygous G/G individuals remained quite low, reaching 5.6 and 12.7 % for C and Cd, respectively (for the distribution according to sex, stage of development and conditions see Fig. S5).

In order to get more insight into the effect of Cd on the genotype of fish, we then calculated the frequency of the G allele according to the condition (C or Cd), stage of development and generation (Fig. 4A). Despite the low number of biological replicates, some

significant differences were observed. In accordance to the differences observed at the methylation level, the G allele frequency was significantly higher in Cd larvae of the F1 and F3 generation in comparison to controls. In addition, considering all individuals (both larvae and adults, n = 18), the frequency of the G allele was significantly higher at the 4th generation (F3) in Cd fish in comparison to controls. Thus, after 3 generations, Cd and C fish were genetically distinct populations for *cep19*. We also noticed a significant difference between Cd adults of the F0 generation and Cd larvae of the F1 generation, i.e. between parents and their offspring. In only one generation, this frequency significantly increased from 0.21 in F0 adults to 0.58 in F1 larvae. Such a result suggests a deviation from the Hardy-Weinberg equilibrium and thus an effect of selection and/or genetic drift. Genetic drift is commonly associated with loss of genetic polymorphism [41]. However, in our case, the percentage of heterozygous individuals increased between the two generations in response to Cd. We must add that a particular effort was made to limit inbreeding. The F1 generation was generated by using a total of 16 (considering that some pairs did not spawn) different mating pairs, these pairs being composed by a male and a female from different families. Concerning selection, no difference was observed in terms of fecundity or reproductive success between Cd and C fish at each generation (the fecundity was however almost significantly lower in Cd fish in comparison to controls at the F1 generation (P = 0.082); Fig. S6). Moreover, no significant difference was observed in mortality rates at each generation between Cd and C animals, both at the embryo-larval and adult stage. In general, the highest mortalities were observed during early development (0-14 dpf, Fig. S6). Observed mortality did not exceed 1% at the juvenile-adult stage but cannibalism was observed until 60 dpf (personal observation). However, we must note that an equal magnitude of mortality in the two conditions does not exclude an effect of selection and selection can induce temporally (i.e. according to the stage of development) dynamic genotype-dependent mortalities [42]. Whatever the mechanisms

involved, Cd favored the rapid increase (in only one generation) of the hypermethylated G allele in the experimental population. Considering all animals analyzed by condition (larvae and adults of the 4 generations, n = 72), the frequency of the G allele was significantly higher with a 2-fold increase (Homogeneity chi-square test, X^2 = 11,47 with a X^2 0.05 = 5.99) in the Cd (0.34) in comparison to the control population (0.17).

Concerning controls, despite no significant variation in the G allele frequency was observed, some fluctuations were observed across generations. The highest frequency (0.33) was observed in adults of the F2 generation. Such a high frequency could explain, at least in part, why no effect of Cd was observed at this generation. The only factor that was not similar among generations was fish density. The density of fish was significantly 2-fold higher in fish of the F2 generation in comparison to the other generations (see Pierron et al., 2021 for details [15]). As presented in Fig. 4B, a significant increase in fish density was associated with a significant increase in the G allele frequency. This suggests that, in addition to Cd, fish density can also affect, to a lesser extent, the proportion of fish presenting the G allele. However, in the present case, high densities did not lead to change in the frequency of the G allele in the subsequent generation (F3).

3.4 Impacts of epi-genetic changes induced by Cd on cep19 transcripts

As methylation of exons can induce alternative splicing events [43], we initially amplified by PCR two regions overlapping the different exons of the *cep19* gene (that comprises 4 exons) using cDNA as template (Fig. S7). If no difference was observed among genotypes for the upstream part (ex1-3) of the mRNA, amplification of the ex3-4 portion was unsuccessful for homozygous A/A individuals (i.e. for the demethylated “version”). We then amplified the ex3-ex4 using genomic DNA as template. Amplicons of different sizes were

obtained for A/A and G/G individuals while heterozygous animals presented the two products (Fig. 5A). After sequencing, we found that the intron 3 (in3) was totally absent in G/G individuals. To confirm our results, all individuals (n = 144) were tested and PCR products were visualized by capillary electrophoresis. This analysis confirmed that the A allele was systematically associated with the in3. On the contrary, when the G base was present, the in3 was absent. We also identified two additional SNPs in the non-coding part (i.e. 3'-UTR) of the ex3 (position 2941886 and 2941888, CGCA→TGTA). Finally, we identified an indel in the 5'UTR region (ex1, Fig. S8). Note that this indel did not induce significant methylation changes in the promoter-ex1 region as no significant difference in methylation levels was observed among genotypes (Fig. S8). With the exception of intron loss, all these changes were already described as they are present in the mRNA sequence of *cep19* (sequence NM_001033734.1). Indeed, the RefSeq transcript (NM_001033734.1) has 4 substitutions, 1 non-frameshifting indel compared to the RefSeq genomic sequence (GRCz11, NC_007133.7, position 2,937,485-2,942,954). To summarize, at the genomic level, individuals with the G/G genotype presented the exonic sequences of the RefSeq transcript and lacked in3. Results are summarized in Fig. 5B. In addition, the two alleles

encode for different transcripts. When the in3 is present (A allele), the ex4 is not transcribed. Through 3'Race on cDNA, we found that the transcript encoded by the A allele retains a portion of the in3 and ends with a region rich in CpG sites (end of the cDNA: base position 2942037 in GRCz11, sequence CGCTCGCGGCATAACAGA). Remarkably, all genetic variations identified between the two alleles involved CpG sites. In this view, we must note that the in3 is fully constituted by a transposable element (TE) called Harbinger-5N1 (also known as DNA-1-4_DR). Generally, TEs are fully methylated to prevent their activation [22]. Accumulating evidences suggest that TEs can act as "methylation centers" after insertion into host DNA, whereby methylation subsequently spreads into the flanking DNA [4]. However,

in our case, the presence of the TE was associated with the demethylation of the ex3, reinforcing the role of the SNP (A/G) in the methylation state of the exon.

We then analyzed the transcription levels of 3 different portions of the *cep19* mRNA corresponding to the ex1, ex3-in3 and ex3-ex4 in larvae (Fig. 5C). As expected from our previous analyses, the amplicon ex3-in3 was not detected in G/G individuals while the amplicon ex3-ex4 was not detected in A/A individuals. We also found that the transcription level of ex3-ex4 was significantly lower in comparison to the transcription level of ex3-in3 in heterozygous animals. This is in accordance with the consensus view that DNA methylation leads to "gene silencing", i.e. transcriptional repression [1]. In this sense, the methylation level of ex1 was significantly lower in fish with a G/G genotype in comparison to other genotypes. Moreover, contrary to A/A individuals, for which no difference was observed between the transcription level of ex1 and ex3-in3, the transcription level of ex3-ex4 was significantly lower than that of ex1 in G/G individuals. This suggests that the methylation of ex3 is not only responsible for the downregulation of *cep19* but also, for some copies, for a precocious arrest in mRNA transcription.

3.5 Impacts of Cd-induced epi-genetic changes on fish growth

In order to investigate a potential link between Cd-induced epi-genetic changes and fish growth, we then calculated the Fulton condition factor of adult fish (124 dpf) according to their genotype and condition. Whereas Cd fish presenting a A/A genotype had a Fulton that was significantly lower than controls, heterozygous Cd fish showed a Fulton that did not differ from C fish (Fig. 6A). In order to deepen and confirm these results, the same analyses were carried out on juvenile fish (35 dpf) of the F0 generation, i.e. in fish for which the weight or condition is not influenced by sex and gonadal weight. The same results were

obtained. Heterozygous Cd fish presented a Fulton that did not differ from controls whereas A/A Cd fish presented a condition that was significantly lower than controls (Fig. 6B). In order to gain additional insight, we then investigated the length (L)-weight (W) relationship (LWR) in adults (Fig. 6C). In fish, the LWR is expressed by using the equation: $W = aL^b$, where b is a growth coefficient used to assess the growth trajectory of fish, indicating whether fish body length increases at the same relative rate as weight. Using the linear regression of the log-transformed equation: $\log(W) = \log(a) + b \log(L)$, the growth coefficient b represents the slope of the relationship. A positive allometric growth (i.e. short and stout) is observed when b is greater than 3.1 and $b < 2.9$ is associated with a negative allometric growth (i.e. long and thin) [44]. This coefficient was inferior to 3 only in the case of Cd A/A fish. Moreover, this coefficient was significantly lower in homozygous A/A Cd fish in comparison to A/A control or to heterozygous A/G Cd fish (data from homozygous G/G fish were not taken into account, see Fig. 6C). Thus, heterozygosity seems to allow Cd fish to reach a growth pattern similar to controls by increasing fish weight gain (rather than decreasing growth in length).

Our results strongly support an adaptive role of the epi-genetic variations and associated changes in relative transcript variants abundance in response to Cd exposure. In Cd fish, heterozygosis was associated with an improved condition. Whereas Cd fish of the F0 and F1 generation presented a condition that was significantly lower than controls (Fig. 2C), epi-genetic changes induced by Cd seem to participate to the rapid recovery of the condition of fish, mainly by increasing their weight gain. In other words, the increase in the frequency of the G allele and concomitantly in the number of heterozygous animals, allowed the Cd population to reach a condition similar to the control one after only two generations.

4. Conclusion

Our results highlight a complex and tight interplay between genetic and epigenetic mechanisms. Mainly, we found that change in one base (SNP) can strongly influence the methylation levels of flanking CpG sites, potentially by affecting the sensitivity of the DNA sequence to TFs. We also found that this CpG-SNP (A/G) was accompanied by several other genetic changes throughout the *cep19* sequence. All these variations were found in the RefSeq transcript. Thus, the SNP was not just a point mutation but rather corresponded to the presence of two different alleles. Cd exposure significantly favored the hypermethylated G allele. In this view, we must underline that the RefSeq transcript record has not yet been subject to final NCBI review. Whereas the RefSeq transcript sequence was derived from the ancient mRNA sequence (BC097138.1) published in 2006, the RefSeq genomic sequence (GRCz11) was published in 2018. This suggests that the emergence of the in3 results from a TE insertion that has arisen recently during zebrafish domestication and/or that the domestication has favored the hypomethylated A allele. This appears to be consistent with the fact that (i) we do not observe transposon footprints in the G allele (i.e. in most cases, when a TE is excised it leaves small rearrangements [45]) and (ii) a loss of genetic polymorphism and a severe reduction in heterozygosity was observed in domesticated zebrafish in the last years [46]. In our case, Cd exposure thus led to an increase in the frequency of a more ancient allele leading to an increased heterozygosity in the experimental population. This “reversion” was associated with an improved condition across generations of fish ancestrally exposed to Cd, suggesting an adaptive process in response to Cd exposure. As the ancient alternative G allele pre-existed in the experimental zebrafish population (at low frequency, reaching 0.17 in controls), our results suggest that standing epi-genetic variation fueled the rapid adaptation of the population to Cd [42,47]. Considering the rapidity and the magnitude of this increase (particularly from the F0 to the F1 generation), the underlying mechanisms require further

investigations. Our work emphasizes the role of the interplay between genetics and epigenetics as well as the exploitation of their standing variation [42,47] in the rapid transgenerational adaptation to anthropogenic environmental stresses.

Acknowledgments

Christian Portier, Kévin Lebaillif and Jacques Martin are thanked for their technical help with the experimental setup.

Data Availability

Raw sequencing data were deposited in the European Nucleotide Archive with accession number PRJEB52137. All other data are included in the article and/or supporting information.

Author Contribution

F.P. conceived the study. F.P. and D.H. performed the experiments. F.P., G.D., F.D., A.B., O.B. and A.R.R. acquired and analyzed the data. F.P. wrote the paper, P.G. and M.N. edited and approved the manuscript.

Funding

This work was supported by the Agence Nationale de la Recherche of France (ANR JCJC TRACE ANR-16-CE34-0008). This Work was supported by the France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by the Agence Nationale pour la Recherche (contrat ANR-10-INBS-09).

References

1. Feil R, Fraga MF. 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13(2):97-109.

2. Allis C, Jenuwein T. 2016. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17:487–500.

3. Sarkies P. 2020. Molecular mechanisms of epigenetic inheritance: Possible evolutionary implications. *Semin. Cell Dev Biol* 97:106-115.

4. Zhou W, Liang G, Molloy PL, Jones PA. 2020. DNA methylation enables transposable element-driven genome expansion. *Proc Natl Acad Sci* USA 117(32):19359-19366.

5. Fitz-James MH, Cavalli G. 2022. Molecular mechanisms of transgenerational epigenetic inheritance. *Nat Rev Genet* 23:325-341.

6. Heard E, Martienssen RA. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157(1):95-109.

7. Horsthemke B. 2018. A critical view on transgenerational epigenetic inheritance in humans. *Nat Commun* 9:2973.

8. Boland MJ, Nazor KL, Loring JF. 2014. Epigenetic regulation of pluripotency and differentiation. *Circ Res* 115(2):311-324.

9. Piferrer F. 2013. Epigenetics of sex determination and gonadogenesis. *Dev. Dyn*. 242(4):360-370.

10. Zeng Y, Chen T. 2019. DNA Methylation Reprogramming during Mammalian Development. *Genes (Basel)* 10(4):257.

11. Head JA. 2014. Patterns of DNA Methylation in Animals: An Ecotoxicological Perspective. *Integr Comp Biol* 54(1):77-86.

12. Skinner MK. 20.14. Environmental stress and epigenetic transgenerational inheritance. *BMC Med* 12:153.

13. Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016. Adaptation to Global Change: A Transposable Element-Epigenetics Perspective. *Trends Ecol Evol* 31(7):514-526.

14. Šrut M. 2021. Ecotoxicological epigenetics in invertebrates: Emerging tool for the evaluation of present and past pollution burden. *Chemosphere* 282:131026.

15. Pierron F, Lorioux S, Heroin D, Daffe G, Etcheverria B, Cachot J, Morin B, Dufour S, Gonzalez P. 2021. Transgenerational epigenetic sex determination: environment experienced by female fish affects offspring sex ratio. *Environ Pollut* 277:116864.

16. Ortega-Recalde O, Day RC, Gemmell NJ, Hore TA. 2019. Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nat Commun* 10:3053.

17. Skvortsova K, Tarbashevich K, Stehling M, Lister R, Irimia M, Raz E. 2019. Bogdanovic, O. Retention of paternal DNA methylome in the developing zebrafish germline. *Nat. Commun* 10:3054.

18. Henson MC, Chedrese PJ. 2004. Endocrine disruption by cadmium, a common environmental toxicant with paradoxical effects on reproduction. *Exp Biol Med* 229(5):383-92.

19. Kelsey G, Feil R. 2013. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Phil Trans R Soc B* 368:1-16.

20. Shalata A, Ramirez MC, Desnick RJ, Priedigkeit N, Buettner C, Lindtner C, Mahroum M, Abdul-Ghani M, Dong F, Arar N et al. 2013. Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *Am J Hum Genet* 93(6):1061-1071.

21. Yıldız Bölükbaşı E, Mumtaz S, Afzal M, Woehlbier U, Malik S, Tolun A. 2018. Homozygous mutation in *CEP19,* a gene mutated in morbid obesity, in Bardet-Biedl syndrome with predominant postaxial polydactyly. *J Med Genet* 55(3):189-197.

22. Pierron F, Daffe G, Lambert P, Couture P, Baudrimont M. 2019. Retrotransposon methylation and activity in wild fish (*A. anguilla*): a matter of size. *Environ Pollut* 245:494-503.

23. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71-W74.

24. McCurley AT, Callard GV. 2008. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. *BMC Mol Biol* 9:102.

25. Ribas L, Piferrer F. 2014. The zebrafish (*Danio rerio*) as a model organism, with emphasis on applications for finfish aquaculture research. *Rev Aquacult* 6:209-240.

26. Yong WS, Hsu FM, Chen PY. 2016. Profiling genome-wide DNA methylation. *Epigenetics & Chromatin* 9:26.

27. Messeguer X, Escudero R, Farré D, Nuñez O, Martínez J, Albà MM. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18(2):333-334.

28. Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, Messeguer, X. 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res* 31(13):3651-3653.

29. Ritter DI, Dong Z, Guo S, Chuang JH. 2012. Transcriptional Enhancers in Protein-Coding Exons of Vertebrate Developmental Genes. *PLoS ONE* 7(5):e35202.

30. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C. 2009. Silent (synonymous) SNPs: should we care about them? *Methods Mol Biol* 578:23-39.

31. Velasco G, Ulveling D, Rondeau S, Marzin P, Unoki M, Cormier-Daire V, Francastel C. 2021. Interplay between Histone and DNA Methylation Seen through Comparative Methylomes in Rare Mendelian Disorders. *Int J Mol Sci* 22(7):3735.

32. Thomassin H, Flavin M, Espinás ML, Grange T. 2001. Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J* 20(8):1974-83.

33. Klengel T, Mehta D, Anacker C. *et al.* 2013. Allele-specific *FKBP5* DNA demethylation mediates gene–childhood trauma interactions. *Nat Neurosci* 16:33–41.

34. Wiechmann T, Röh S, Sauer S. *et al.* 2019. Identification of dynamic glucocorticoid-induced methylation changes at the *FKBP5* locus. *Clin Epigenet* 11:83.

35. Kremsky I, Corces VG. 2020. Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. *Genome Biol* 21(1):118.

36. Wilson KS, Tucker CS, Al-Dujaili EAS, Holmes MC, Hadoke PWF, Kenyon CJ, Denvir MA. 2016. Early-life glucocorticoids programme behaviour and metabolism in adulthood in zebrafish. *J Endocrinol* 230(1):125-142.

37. Faught E, Vijayan MM. 2019. Loss of the glucocorticoid receptor in zebrafish improves muscle glucose availability and increases growth. *Am J Physiol Endocrinol Metab* 316(6):E1093-E1104.

38. Jackson SE, Kirschbaum C, Steptoe A. 2017. Hair cortisol and adiposity in a population-based sample of 2,527 men and women aged 54 to 87 years. *Obesity* 25:539-544.

39. Björntorp P, Rosmond R. 2000. Obesity and cortisol. *Nutrition* 16(10):924-936.

40. Ronco AM, Urrutia M, Montenegro M, Llanos MN. 2009. Cadmium exposure during pregnancy reduces birth weight and increases maternal and foetal glucocorticoids. *Toxicol Lett* 188(3):186-191.

41. Masel J. 2011. Genetic drift. *Curr Biol* 21(20):R837-R838.

42. Bitter MC, Kapsenberg L., Gattuso JP, Pfister CA. 2019. Standing genetic variation fuels rapid adaptation to ocean acidification. *Nat Commun* 10:5821.

43. Shayevitch R, Askayo D, Keydar I, Ast G. 2018. The importance of DNA methylation of exons on alternative splicing. *RNA* 24(10):1351-1362.

44. Forster CE, Norcross BL, Spies I. 2020. Documenting growth parameters and age in Arctic fish species in the Chukchi and Beaufort seas. *Deep Sea Res Part II: Top Stud Oceanogr* 177: 104779.

45. Muñoz-López M, García-Pérez JL. 2010. DNA transposons: nature and applications in genomics. *Curr Genomics* 11(2):115-128.

46. Suurväli J, Whiteley AR, Zheng Y, Gharbi K, Leptin M, Wiehe T. 2020. The Laboratory Domestication of Zebrafish: From Diverse Populations to Inbred Substrains. *Mol Biol Evol* 37(4):1056–1069.

47. Lai YT, Yeung CKL, Omland KE, Pang EL, Hao Y, Liao BY, Cao HF, Zhang BW, Yeh CF, Hung CM, Hung HY, Yang MY, Liang W, Hsu YC, Yao CT, Dong L, Lin K, Li SH. 2019. Standing genetic variation as the predominant source for adaptation of a songbird. *Proc Natl Acad Sci USA* 116(6):2152-2157.

Figure 1

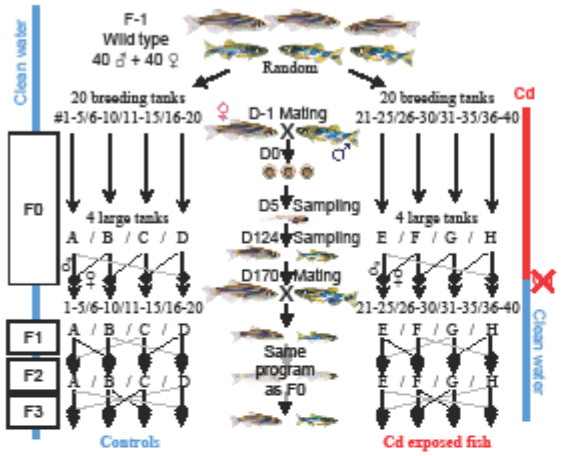


Figure 2

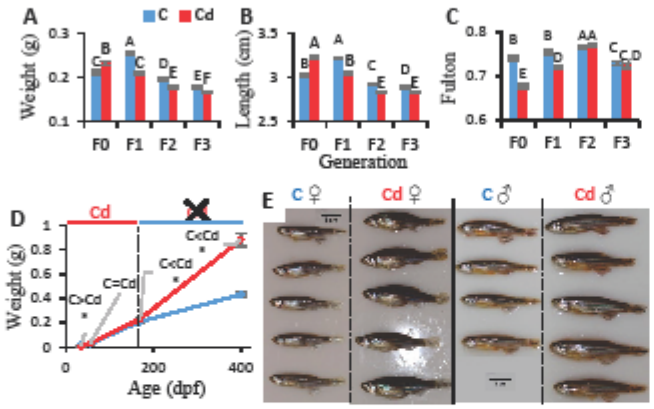


Figure 3

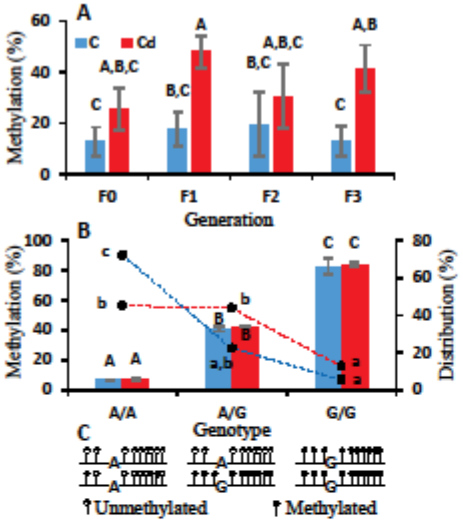


Figure 4

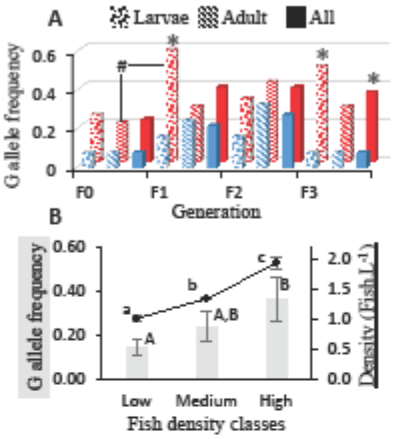


Figure 5

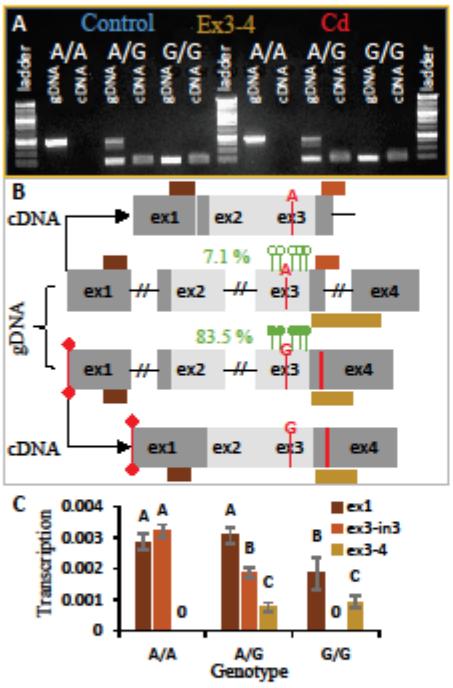
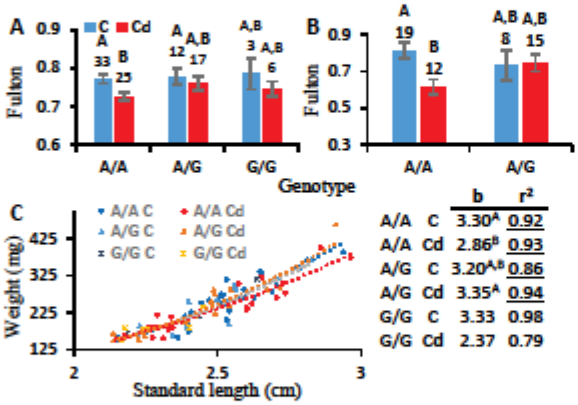


Figure 6



		b	r^2
A/A	C	3.30^{A}	0.92
A/A	Cd	2.86^{B}	0.93
A/G	C	$3.20^{A,B}$	0.86
A/G	Cd	3.35^{A}	0.94
G/G	C	3.33	0.98
G/G	Cd	2.37	0.79