



HAL
open science

Population transcriptogenomics highlights impaired metabolism and small population sizes in tree frogs living in the Chernobyl Exclusion Zone

Clément Car, André Gilles, Elen Goujon, Marie Laure Delignette-Muller, Luc Camoin, Sandrine Frelon, Pablo Burraco, Samuel Granjeaud, Emilie Baudelet, Stéphane Audebert, et al.

► To cite this version:

Clément Car, André Gilles, Elen Goujon, Marie Laure Delignette-Muller, Luc Camoin, et al.. Population transcriptogenomics highlights impaired metabolism and small population sizes in tree frogs living in the Chernobyl Exclusion Zone. *BMC Biology*, 2023, 21 (1), pp.164. 10.1186/s12915-023-01659-2 . irsn-04300523

HAL Id: irsn-04300523

<https://asnr.hal.science/irsn-04300523v1>

Submitted on 23 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



Population transcriptogenomics highlights impaired metabolism and small population sizes in tree frogs living in the Chernobyl Exclusion Zone

Clément Car^{1,2}, André Gilles^{3*}, Elen Goujon^{1,2,4}, Marie-Laure Delignette Muller⁵, Luc Camoin⁶, Sandrine Frelon^{1,2}, Pablo Burraco^{7,8}, Samuel Granjeaud⁶, Emilie Baudalet⁶, Stéphane Audebert⁶, Germán Orizaola^{7,9,10}, Jean Armengaud¹¹, Arthur Tenenhaus⁴, Imène Garali^{1,2}, Jean-Marc Bonzom^{1,2} and Olivier Armant^{1,2*} 

Abstract

Background Individual functional modifications shape the ability of wildlife populations to cope with anthropogenic environmental changes. But instead of adaptive response, human-altered environments can generate a succession of deleterious functional changes leading to the extinction of the population. To study how persistent anthropogenic changes impacted local species' population status, we characterised population structure, genetic diversity and individual response of gene expression in the tree frog *Hyla orientalis* along a gradient of radioactive contamination around the Chernobyl nuclear power plant.

Results We detected lower effective population size in populations most exposed to ionizing radiation in the Chernobyl Exclusion Zone that is not compensated by migrations from surrounding areas. We also highlighted a decreased body condition of frogs living in the most contaminated area, a distinctive transcriptomics signature and stop-gained mutations in genes involved in energy metabolism. While the association with dose will remain correlational until further experiments, a body of evidence suggests the direct or indirect involvement of radiation exposure in these changes.

Conclusions Despite ongoing migration and lower total dose rates absorbed than at the time of the accident, our results demonstrate that *Hyla orientalis* specimens living in the Chernobyl Exclusion Zone are still undergoing deleterious changes, emphasizing the long-term impacts of the nuclear disaster.

Keywords Genetics, Transcriptomics, Population, Amphibian, Pollution, Ionizing radiation, Chronic exposure, Modelling, SNP

*Correspondence:

André Gilles
andre.gilles@univ-amu.fr
Olivier Armant
olivier.armant@irsn.fr

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Rapid environmental change caused by human activities poses threats to many wild local species by increasing their vulnerability and notably increasing the risk of population extinction. Individual functional modifications of biological features (i.e. from gene to organism) determine this vulnerability [1]. Here, “function” is referring to a causal relationship between environmental trigger and individual response which impact the interaction between individuals and their environment [2]. These modifications can be detrimental, neutral or adaptive, as well as transmitted or non-transmitted. Functional adaptations in response to an altered environment have often been studied to disentangle plastic and selected processes [3–5]. But the interaction between evolutionary processes (irrespective of functional consequences, and adaptive nature) and plastic changes in the emergence of such functional modifications remains largely an open issue.

The current status of species and their ability to persist locally can be first explored by examining the evolutionary processes faced with contemporary environmental change [6]. By investigating population genetic diversity and structure, population genetics contributes towards disentangling the mechanisms that shape these evolutionary processes [7]. Short-term selection and changes of evolutionary potential in response to new abiotic and biotic interactions induced by human activities have often been emphasised for example in the case of urbanisation [8], climate change [9], or pollution [10]. This generally leads to a decrease in genetic diversity on standing variation or by selective sweeps. Other processes like migration can be also greatly modified by human activities and influence gene flow between populations, potentially increasing genetic diversity with population admixture [11] or decreasing genetic diversity by altering connectivity [12]. More broadly, in the case of small population size, random fixation of transmitted variations through genetic drift can lead to decrease in genetic diversity which can in turn affect population vulnerability [13, 14]. Also, anthropogenic abiotic changes like pollution can induce *de novo* mutations which could lead to an increased genetic diversity [15], thereby providing the source of genetic variability upon which evolution acts, but may have deleterious effects by directly impairing protein function [16, 17] or through indirect collateral effects [10, 18, 19]. In parallel with evolutionary processes, phenotypic plasticity interacts with individual functional changes and notably influences adaptive evolution. In small populations the increase of genetic drift limits the genetic potential on which selection can act and thus increases the importance of plastic changes to be able to cope with the altered environment [20]. On the

other side, individual functional changes act also as feedback on evolutionary processes. For example, if plastic changes lead to a phenotype close to the fitness optimum, directional selection may only have a limited effect [20]. Combining the study of evolutionary processes (with or without direct individual functional consequences) with more functional-centred effects appears thus essential to estimate properly the vulnerability of populations facing anthropogenic pressures.

Pollutants constitute an interesting case study to analyse the evolutionary processes, phenotypic plasticity and functional responses of exposed organisms as they are examples of rapid and persistent environmental changes [21, 22]. In particular, radioactive contaminations after major nuclear accidents like the Chernobyl accident are characterised by a wide range of consequences on wildlife [23–25] and long environmental persistence [26]. The short-term effects of high levels of radionuclides on wild organisms are relatively well described [23]. But even if the regulation in the most contaminated places (as for example in Ukraine with the establishment of the Chernobyl Exclusion Zone greatly limiting human activities [27, 28]) promotes the presence of a large number of species (including large mammals [29]), we are still unable to describe the long-term vulnerability of populations and ecosystems [30, 31]. A good example is the bank voles (*Myodes glareolus*) living in the Chernobyl Exclusion Zone (CEZ), for which a large number of studies showed that they are still experiencing individual functional modifications along the radioactive gradient more than 30 years after the nuclear power plant accident (e.g. colouration [32], cataracts [33], organ mass [34], DNA damage response [35]). These modifications were associated with decreased populations size [36] and thus question the impact on the vulnerability of populations in contaminated places.

Here, we take the Chernobyl study case to assess the long-term consequences of such a major anthropogenic environmental change and the nature of these consequences. In particular, we attempted to conciliate the question of evolutionary processes with a more individual functionally centred study. We focused on the Eastern tree frog *Hyla orientalis*, a species of particular interest because of its relatively small dispersal radius (about 0.5 km per generation [37]) and its exposure to radionuclides in different components of the environment as it lives buried in the ground during winter, reproduces and develops in water and lives as adult in the vegetation [38]. These characteristics make this species relevant to characterise the effects of ionizing radiation (IR) and inform on the possible non-perceptible effects on other components of the biotic community, as proposed for ‘sentinel species’ [39, 40]. While over 3 years (from 2016 to 2018)

the total dose rate absorbed by male tree frogs during the breeding period (ranged from 0.07 to 32.40 $\mu\text{Sv/h}$) [41] was estimated to be below the no-effects threshold values (i.e. < to 42 $\mu\text{Gy/h}$) determined by the International Commission on Radiological Protection (ICRP), a recent population genetics study indicated a singular evolution of CEZ populations since the accident compared to other European populations [42]. Analyses of the specific mitochondrial profile of tree frogs in the CEZ revealed a 100-fold higher rate of mitochondrial substitution and smaller effective population sizes compared to other European populations. In the same time, while several blood parameters were not shown to be dependent on total dose rate [43], the high mitochondrial substitution rate observed in the CEZ [42] pointed towards potential metabolic impairments.

The power of transcriptomic analysis, and especially the study of gene expression level, to explore the potential effects of pollutants on individual function has already been emphasised [44–46]. But transcriptomic analysis appears also particularly interesting to explore the potential effects of pollutants by assessing systemic evolutionary processes in the transcribed regions by population genetics analysis and estimate functional effects via the study of gene expression [47–49]. In this study, the de novo transcriptome assembly of *H. orientalis* allowed the comprehensive transcriptomics analysis of 87 individuals sampled in 2018 across a gradient of dose rates in the CEZ. The molecular changes potentially induced by radioactive contamination were assessed by dose–response modelling with specific emphasis on coregulation of genetic networks with environmental variables, and including the body condition index, which was used as an integrative measure of major physiological changes at the individual scale [50]. Expressed single-nucleotide polymorphisms (SNP) derived from transcriptomics data were used to analyse the population structure and genetic variation within the Chernobyl area and to assess the intensity of evolutionary processes ongoing inside the CEZ for this species.

Results

Negative correlation between body condition and individual dose of ionizing radiation

To better estimate the current exposure to IR in wild tree frogs, 87 breeding Eastern males tree frog (*Hyla orientalis*) were collected in the CEZ and in non-radiocontaminated area near Slavutych (Fig. 1a and Additional file 1: Table S1) and their individual total dose rate (ITDR) (Fig. 1b) and body condition index (BCI) were estimated. A statistically significant decrease in BCI along the contamination gradient was observed (Chisq=4.48, $p=0.034$), due to a higher frequency of individuals with

BCI below 0 in highly contaminated sites compared to other sites (chi-square test $p=0.0003$, 61% $n=43$ in sites A18, C18 and D18 compared to 21% $n=44$ in sites B18, E18, F18, G18, H18) (Fig. 1c).

Geographically structured population and asymmetrical admixture in the CEZ

The de novo transcriptome assembly of *Hyla orientalis* was performed by using the combined set of 5.9 billion RNASeq paired end-reads from 5 tissues (tibia muscle, heart, eye, brain and testis) obtained from three specimens collected near the Slavutych locality. The E90N50 of the assembly was 1690 bp, and 95% of the conserved genes in Tetrapoda were completely assembled (BUSCO) [51]. The reference map of the *Hyla orientalis* transcriptome was defined as a set of 17,323 non-redundant assembled contigs. From the 7286 contigs encoding peptides detected by RNAseq in the muscle tissue (Tag Per Million, TPM>0.1), 3611 (50%) corresponding proteins were also detected by proteomics (Additional file 2: Fig. S1 and Additional file 3: Supplementary Material and Methods).

We then generated RNAseq data from tibia muscle of 87 individuals sampled in the CEZ and Slavutych region, with similar coverage (Additional file 4: Table S2, Additional file 2: Fig. S2a, and Additional file 3: Supplementary Material and Methods). The tibia muscle was dissected from one leg for all individuals as it is an abundant and easily accessible tissue in frogs, deeply innervated and with high energy demand for locomotion [52]. From these data, we collected 374,102 variants and selected 157,677 SNPs (42.15%) based on stringent quality control criteria to assess genetic diversity. The ratio of nucleotides transition/transversion (Ti/Tv) was between 2.0 and 2.2 for all individuals, as expected for this type of data [53]. To estimate the intensity of gene flow within the sampled area, and also to make an initial assessment of the relevance of the markers used, we first examined the population genetic structure. We detected that all populations inside and outside the CEZ were moderately differentiated with $G'st$ comprised between 0.054 and 0.115. The genetic distances ($G'st/(1-G'st)$) were significantly and positively correlated to geographical distances, indicative of an isolation by distance (Mantel, $r=0.6581$, $p=0.011$). This isolation by distance remained significant even by controlling for dose rate differences (partial Mantel, $r=0.5915$, $p=0.012$). Thus, genetic differentiation is explained on average, by affiliation reflected by geographical distances, and not only by the averaged total dose rates (ATDR) (see Materials and Methods for more information about the ATDR). To further describe population genetic structure in the CEZ, we conducted a principal component analysis (PCA) using bi-allelic SNPs.

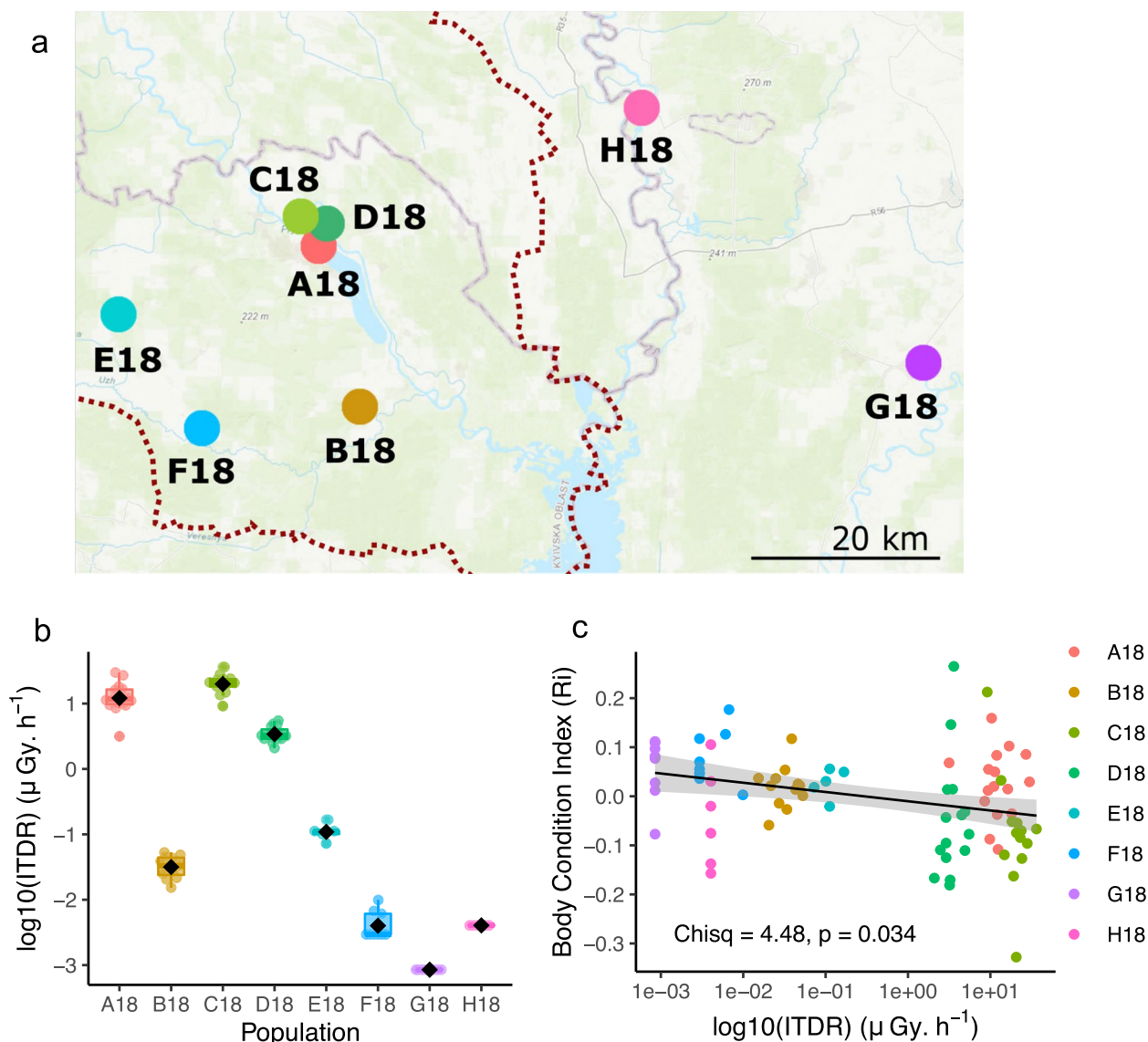


Fig. 1 Geographical distribution, individual dose rate (ITDR) and body condition of sampled frogs. **a** Map showing the sampling sites (letters) in the Chernobyl Exclusion Zone (dashed line) and the Slavutych region. **b** Distribution of the ITDR ($\mu\text{Gy}\cdot\text{h}^{-1}$) in the different sampled populations. **c** Relationship between ITDR and BCI (R_i). The linear model is represented by a black line with 95% confidence intervals

The Slavutych populations (G18 and H18) were well separated from CEZ populations (Additional file 2: Fig. S2b), but only a relatively low variation was explained by the two first principal components (PC) (5.3% for PC1 and 3.66% for PC2). To further infer relationships between genetic clusters, we carried out a discriminant analysis of principal components (DAPC) in which populations were implemented a priori. The populations were clearly distributed according to geography, with Slavutych populations (G18 and H18) more differentiated from CEZ populations, as expected (Fig. 2a). The genetic structure

of CEZ populations was also consistent with their geography. Indeed, the two southern populations B18 and F18 were close together and distant from the three central populations A18, C18 and D18, the south-west population E18 bridging these two groups. We then assessed the probability of affiliation of individuals to each sampled population based on their genotype (Fig. 2b). Individuals displayed high membership a priori with little to moderate mixing between them, with the exception of the two Slavutych populations that were completely separated from CEZ populations. The

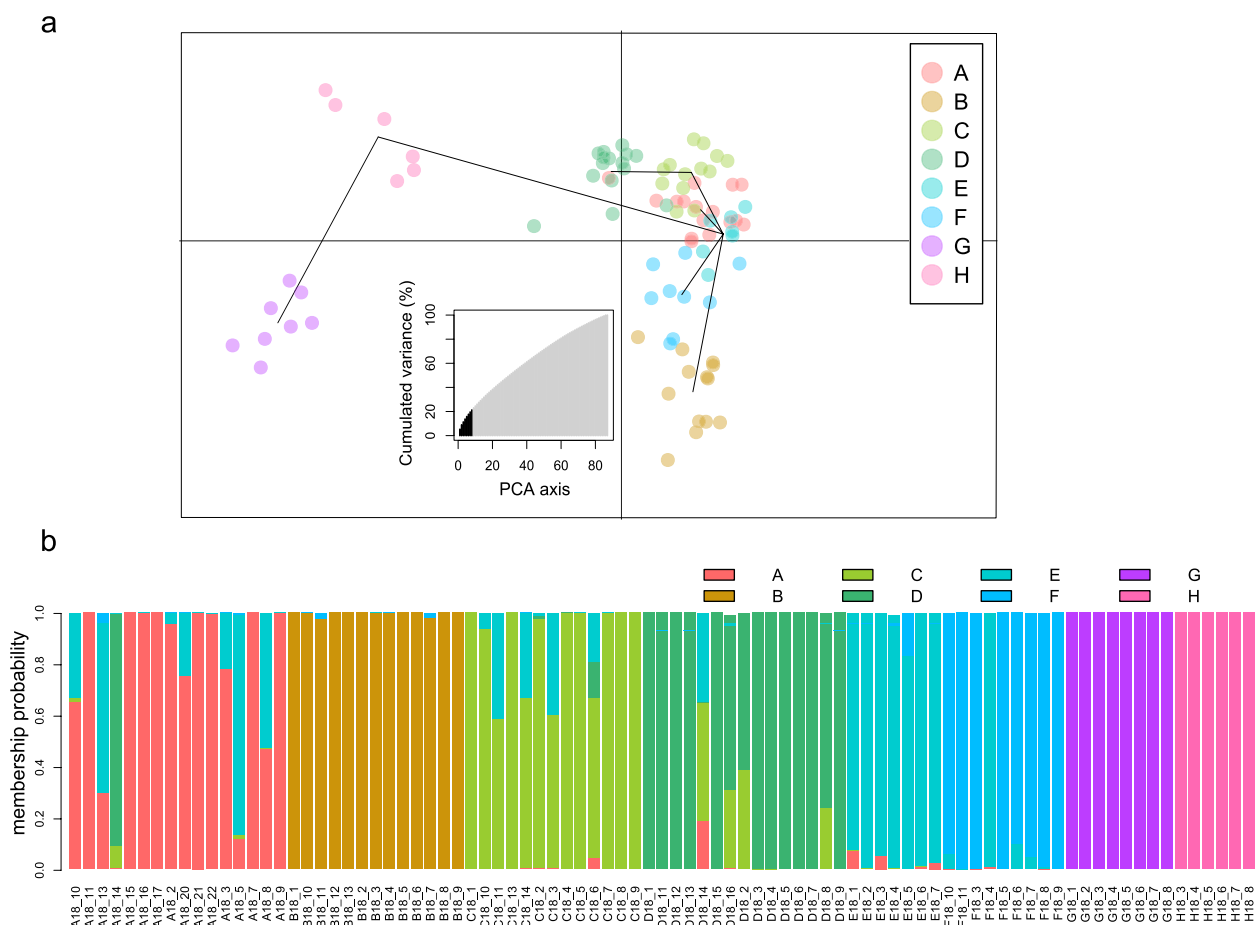


Fig. 2 Population structure of *Hyla orientalis* in the Chernobyl region based on genetic polymorphism. **a** DAPC of 87 individuals. **b** Probability of affiliation (membership) of each individual to each sampling site

three central populations in the CEZ, A18, C18 and D18, presented the highest number of admixed individuals, reflecting their geographical proximity. We also noted a higher proportion of admixed individuals coming from the south (E18 and F18 populations) towards the centre of the CEZ but not the converse (Fig. 2b). These observations were confirmed by non-supervised *k*-means clustering of genetic polymorphisms and by admixture analysis based on ancestry coefficient inferences, without a priori on population affiliation (Additional file 3: Fig. S2c and d).

Functional prediction of SNP outliers correlated with individual total dose rate (ITDR)

We next aimed at describing the potential selective process ongoing in the CEZ by testing loci associated to the dose gradient and including the confounding effect due to population structure. We detected 21,432 SNPs highly correlated to ITDR (q -value < 0.05), scattered

among expressed exons of 5283 different contigs encoding functionally characterised orthologous genes. A total of 4138 SNPs non-synonymous variants were detected and associated with genes involved in ATPase activity (GO:0016887 q -value < 10^{-3}) and small GTPase mediated signal transduction (GO:0051056 q -value < 10^{-2}). Finally, we found that 159 SNPs in 143 different contigs were stop-gained variants predicted to have a functional impact on orthologous proteins with function in energy metabolism including IDH3G (Uniprot P51553), ALDH2 (Uniprot P05091), and LDHA (Uniprot P00338), as well as the stress factors AHSA1 (Uniprot O95433) and EIF2AK3 (Uniprot Q9NZJ5) (Additional file 5: Table S3). The number of homozygous stop-gained mutations deviating from Hardy–Weinberg (HW) equilibrium was higher in populations highly contaminated in the CEZ compared to the others (exact test of HW equilibrium p < 0.01, n = 8 in sites A18, C18, D18, n = 4 in sites B18, E18, F18 and n = 0 for G18 and H18) (Additional file 6: Table S4).

Negative relationship between genetic diversity indices and the averaged total dose rates (ATDR)

The nucleotide diversity (π) of populations located at the centre of the CEZ (A18, C18 and especially D18) was lower than the one observed at the south of the CEZ (B18, E18, F18) and in the Slavutych region (G18 and H18) (Fig. 3a). Furthermore, π and individual heterozygosity were both negatively correlated with the ATDR (respectively, $S=148$, $\rho=-0.760$, $p=0.037$; $F=9.445$, $p=0.003$) (Fig. 3a, b). As decrease of heterozygosity could be a sign of increased inbreeding within the populations in the CEZ, we assessed the inbreeding coefficient (Fis) for each population. Average Fis by population

was comprised between $-0.130 < Fis < -0.026$, indicating a general excess of heterozygosity. Higher Fis indices were observed in populations highly contaminated in the CEZ (Fig. 3c) and Fis values were positively correlated to ATDR ($S=12$, $\rho=0.857$, $p=0.011$) even including only the loci at HW equilibrium ($S=10$, $\rho=0.881$, $p=0.007$). Thus, the populations studied here do not show general sign of inbreeding, despite a decrease in genetic diversity in most contaminated populations. We next assessed variability of relatedness inferred by pairwise kinship coefficient between sites. Higher variations in relatedness were observed within highly contaminated populations in the CEZ (Fig. 3d) ($z=5.73$, $\tau=0.253$,

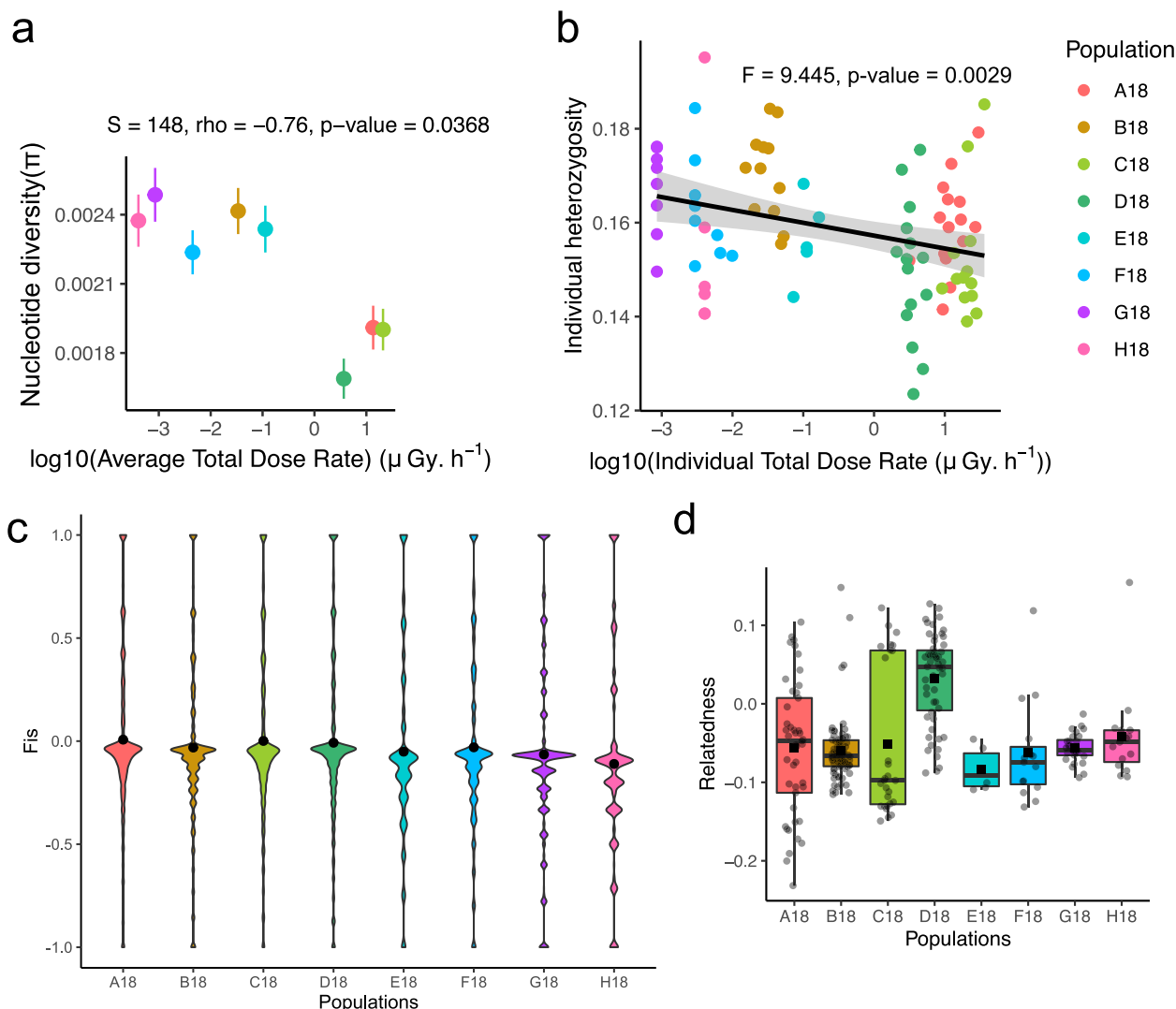


Fig. 3 Population genetics summary and relatedness of *Hyla orientalis* in the Chernobyl region. **a** Correlation plot of nucleotide diversity (π) on averaged total dose rate (ATDR) represented at the mean and standard error per sampled population. **b** Individual heterozygosity on individual total dose rate (ITDR). The linear model is represented by a black line with 95% confidence intervals. **c** Representation of the distribution of individual heterozygosity. **d** Degree of relatedness per population

$p=9.93.10^{-9}$) with notably a higher percentage of individuals with relatedness higher than zero in the three most exposed sites (15%) than in lower exposed sites (2%) (Chisq=41.26, $p=1.33.10^{-10}$). Together, these results indicate that individuals sampled in highly contaminated area in the CEZ do not show sign of higher inbreeding but display higher degree of relatedness.

Dose dependent effects in the CEZ assessed by transcriptomics clustering analysis

Individuals were discriminated in two main groups by hierarchical clustering of contig expression (Additional file 2: Fig. S3). To test how ITDR contributes to this observed pattern, we tested the correlation between individual cophenetic distances (as an estimate of similarity between individual) and ITDR. We found a significant positive correlation between cophenetic distances and ITDR (Mantel, $r=0.121$, $p=1.10^{-4}$). Importantly, this correlation remained significant even by controlling for genetic distances (partial Mantel, $r=0.126$, $p=1.10^{-4}$) or geographical distances (partial Mantel, $r=0.101$, $p=2.10^{-4}$). Thus, these results highlight that changes in expression patterns observed by transcriptomics analysis are correlated to the ITDR, and importantly this correlation does not depend only on geography or genetics.

Impact of long-term exposure to radioccontaminants in the CEZ by transcriptomic analysis

The differential expression analysis of 17,323 contigs was performed using individuals sampled outside the CEZ as reference. Comparison of fold changes was correlated ($\rho>0.63$) using either reference populations (G18 or H18) or a lowly contaminated site (F18) showing that the dose effects were prominent, and choice of reference sites did not impair the identification of differentially expressed contigs (Additional file 2: Fig. S4). Using the most distant site (G18) as reference, we found 477 differentially expressed genes (DEG) in common in the pair-wise comparative analysis made for the highly contaminated sites (A18, C18, D18) and 367 DEG common for the lowly contaminated sites (B18, E18, F18). The differentially expressed contigs were involved in pathways involved in oxidoreductase activity (GO:0016705,

$q\text{-value}<1.10^{-2}$) and glycolysis and gluconeogenesis (hsa00010, $q\text{-value}<1.10^{-2}$) (Additional file 7: Table S5).

We then analysed the transcriptomics changes potentially associated with the environmental changes in the 87 individuals. Several groups of contigs were characterised by relative higher expression in individuals sampled mainly from sites A18, C18 and D18 (clusters 1 and 2), while a second group of contigs was expressed at low levels in individuals from sampling sites A18, C18 and D18 (cluster 3,4 and 6) (Fig. 4a). These results show that individuals sampled in highly contaminated sites inside the CEZ have a specific transcriptome signature that discriminates them from individuals originating from lowly contaminated sites in the CEZ or outside the CEZ. We then assessed which biological pathways were deregulated in individuals sampled in the CEZ by analysing the GO biological processes, GO molecular functions and KEGG pathways enriched in the different clusters obtained after hierarchical clustering. Contigs present in the cluster characterised by higher levels of expression in the highly contaminated sites were enriched (among others) in aerobic respiration (GO:0009060, $q\text{-value}<10^{-6}$), glycolysis and gluconeogenesis (hsa00010, $q\text{-value}<10^{-5}$) and muscle cell development (GO:0055001, $q\text{-value}=0.0016$). Contigs lowly expressed in highly contaminated individuals in the CEZ were involved in immune response (GO:0002252, $q\text{-value}<10^{-8}$) and tumor necrosis factor signalling (GO:0032640, $q\text{-value}<10^{-2}$) (Fig. 4b, c). Importantly, adjusting contig expression by the genetic distances led to clusters of contig expression with similar GO term composition (Additional file 2: Fig. S5). A complete list of the biological pathways enriched in each cluster is provided in Additional file 8: Table S6. Gene set enrichment analysis was then performed on the different sets of differentially expressed genes (using G18 as a reference, as before). The differentially expressed genes from highly contaminated sites (A18, C18 and D18) were enriched in proteins involved in oxidoreduction processes, while differentially expressed genes in the lowly contaminated sites (B18, E18 and F18) were enriched in proteins with function in lipid metabolism (phosphatidylcholine–sterol O–acyltransferase activator activity). Signalling transducer molecules and structural

(See figure on next page.)

Fig. 4 Analysis of gene expression by transcriptomics analysis in the 87 individuals. **a** Unsupervised hierarchical clustering of normalised expression of 5735 contigs differentially expressed in at least one population (using G18 individuals as control). Top dendrogram corresponds to the clustering of individuals based on overall contig expression, highlighted as two main clusters (black and blue). The expression levels of contigs (as Row Z-score) are represented as follows: low expression (blue), moderate expression (white) and high expression (red). **b** Enrichments of Gene Ontology Biological Process and **c** KEGG in the 9 different clusters obtained by hierarchical clustering. The numbers in brackets correspond to the total number of contigs in each cluster (bottom, x-axis). **d** Treeplot based on semantic similarity of GO molecular function obtained after by gene set enrichment analysis (GSEA). The different biological pathways enriched in each gene set (comparative analysis between each site and using G18 as a reference) are indicated at the bottom of the heatmap. The colour-code from blue to red depicts the $q\text{-value}$ of the enriched GO terms (grey colour indicates no enrichment)

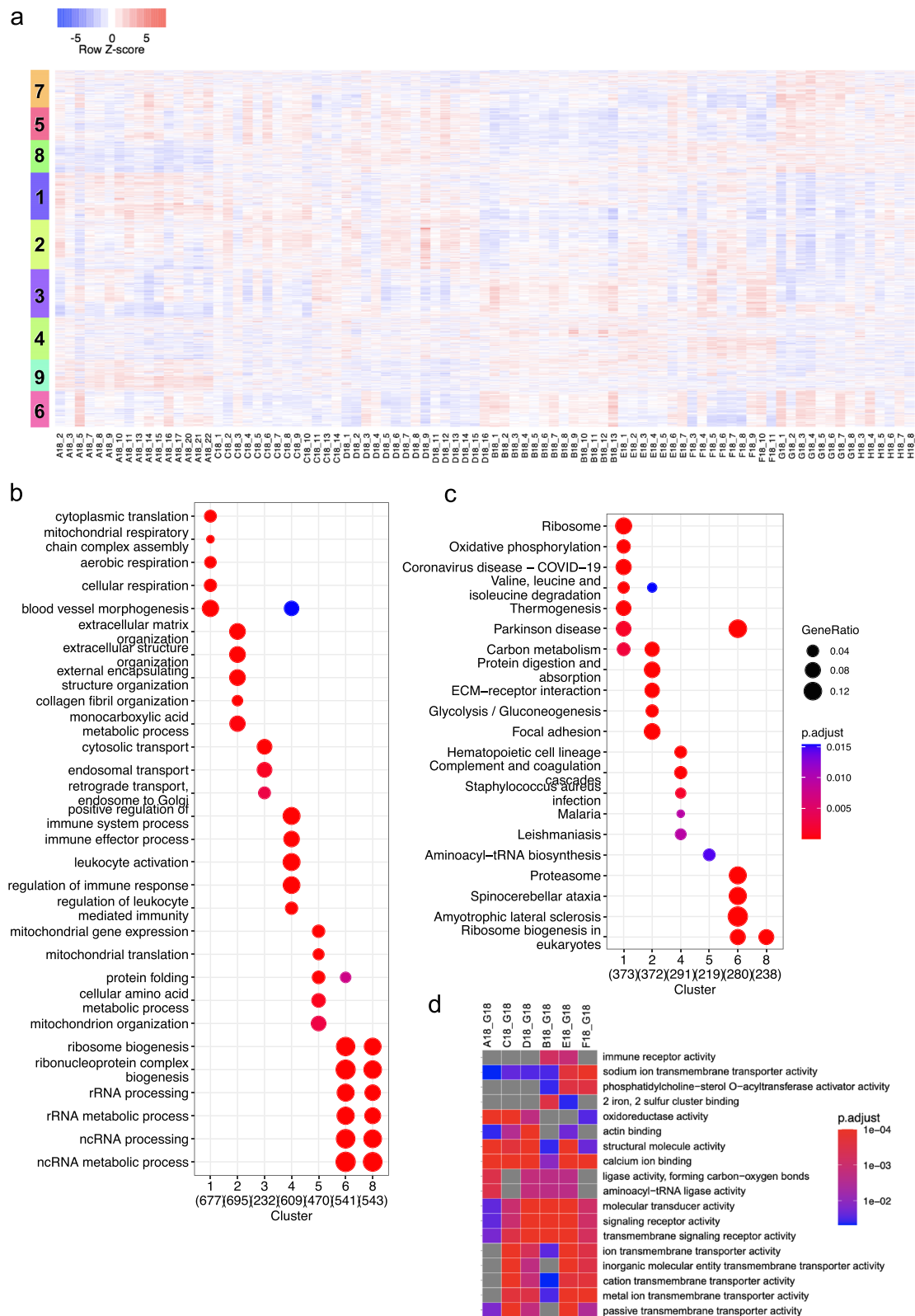


Fig. 4 (See legend on previous page.)

components of actin cytoskeleton were enriched in all gene sets (Fig. 4d).

We then modelled contigs expression profiles along the dose-gradient [54] and computed benchmark dose (BMD_{1SD}) to rank the sensitivity of the deregulated GO pathways (Fig. 5a). The subset of contigs with dose-response models were notably involved in the response to oxidative stress (GO:0006979, q -value=0.016), cellular response to lipid (GO:0071396, q -value=0.012) and muscle system process (GO:0003012, q -value=0.042), thereby showing a deregulation of contigs involved in these biological pathways in function of the ITDR. The complete list of the deregulated pathways is provided in Additional file 9: Table S7. Ranking of the BMD_{1SD} highlighted that the pathway involved in reduction of oxidative stress was deregulated at lower ITDR compared to contigs related to muscle contraction ($BMD_{1SD} < 10^{-2} \mu Gy.h^{-1}$ and about $1 \mu Gy.h^{-1}$ respectively) (Fig. 5b, c).

We then applied a non-supervised machine learning methods based on weighted gene co-expression network analysis [55] to identify modules of co-expressed contigs associated with variations of the BCI and ITDR (Fig. 6a). One module was significantly associated with ITDR and enriched in contigs with function in mitochondrial ATP synthesis coupled electron transport (q -value $< 10^{-19}$) and aerobic respiration (q -value $< 10^{-27}$). One other module was associated with the body condition index (BCI) and contained contigs involved in extracellular matrix structural constituent (q -value $< 10^{-39}$) (Fig. 6b and Additional file 10: Table S8). The correlation between the module eigen-gene and ITDR remained significant after controlling for genetic distances (p -value = 0.0017, see Additional file 3: Supplementary Material and Methods). Finally, modules correlated to ITDR or BIC after correcting contigs expression by genetic distance were enriched in ATP synthesis (q -value $< 10^{-4}$) and oxidative phosphorylation (q -value $< 10^{-4}$) (Additional file 2: Fig. S6 and Additional file 11: Table S9). Together, these results demonstrate that the transcriptional regulation of genes involved in several biological processes like ATPase activity, hexose

metabolism and muscle function are affected by IR, and these changes are correlated to a decreased BCI.

Discussion

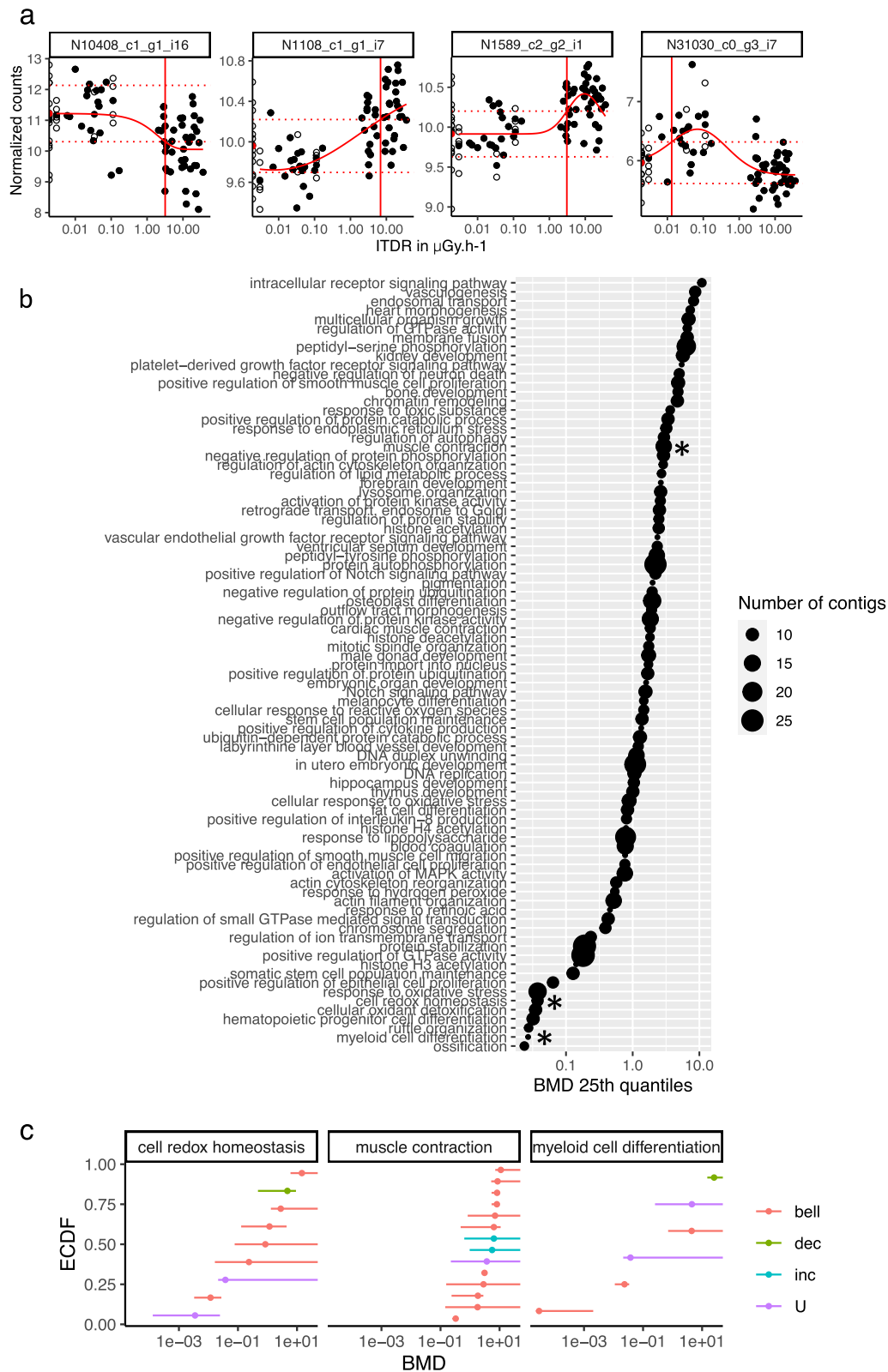
Tree frog (*Hyla orientalis*) populations living in highly contaminated sites inside the CEZ showed lower nuclear genetic diversity and higher relatedness, as a result of small population sizes. In addition to the underlying genetic drift process, we highlighted stop-gained mutations and changes in transcriptional profile strongly associated with the individual dose received which are predicted to have or support functional consequences in genes involved in energetic metabolism, and in turn could explain phenotypic changes including body condition index.

The analysis of genetic variants showed a significant negative correlation of average nucleotide diversity and average individual heterozygosity with ITDR indicative of genetic drift inside the CEZ, likely due to small effective population sizes. Decreased population sizes associated with the contamination of the environment is also corroborated by the higher propensity of sampled individuals to be more related in the most contaminated sites. Our data on inbreeding coefficient did not reveal signs of higher consanguinity inside the CEZ at the level of the genome but rather high variability in relatedness with notably higher degree of relatedness in the most contaminated sites.

By analysing genetic population structure, we highlighted that populations inside and outside the CEZ were genetically differentiated. The Slavutych populations were more differentiated compared to the ones collected inside the CEZ, pointing towards genetic specificity of the CEZ populations even if the geographical distance between Slavutych and CEZ populations is low (36 km for the closest and 66 km on average). The presence of higher number of admixed individuals in the centre of the CEZ indicates a directed gene flow towards the most contaminated places which can reflect past or ongoing migrations from low towards higher contaminated sites, either as a consequence of recolonisation of declined frog populations or the decrease of anthropogenic activities in the centre of the CEZ, as proposed for barn swallows

(See figure on next page.)

Fig. 5 Dose-response modelling and derivation of benchmark mark dose (BMD_{1SD}) from the transcriptomics data. **a** Examples of dose-response modelling for several contigs showing the expression values (black dots) along the dose gradient. The dose-response model computed (red line) is used to determine the BMD_{1SD} (vertical line), defined as the minimal dose for which the signal reach the benchmark response ($BMR = y_0 \pm SD$) with the modelled value at null dose and the SD the estimated residual standard deviation around the model (dotted horizontal lines). The ITDR (in $\mu Gy.h^{-1}$) as a log scale on the x-axis, responses at the control sites (G18 and H18) appear as half circles on the y-axis. **b** Distribution of the enriched GO biological process summarised by the 25th percentile of the BMD_{1SD} values from all modelled contigs belonging to the considered GO term. The size of each point codes for the number of modelled contigs in each GO term. Asterisk symbol (*) indicates the 3 GO term for which empirical cumulative distribution function (ECDF) is shown as example in **c**. **c** Example of ECDF plots used to estimate BMD_{1SD} shown in **b** (*). The 95% confidence intervals are indicated for each selected contigs categorised in 3 different GO biological processes. The dose (in $\mu Gy.h^{-1}$) is indicated on the x-axis. The trend of the dose-response curves fitted for each contig is indicated by the colour code (bell: bell-shape, U: U-shape, dec: decreasing (downregulated), inc: increasing (upregulated))



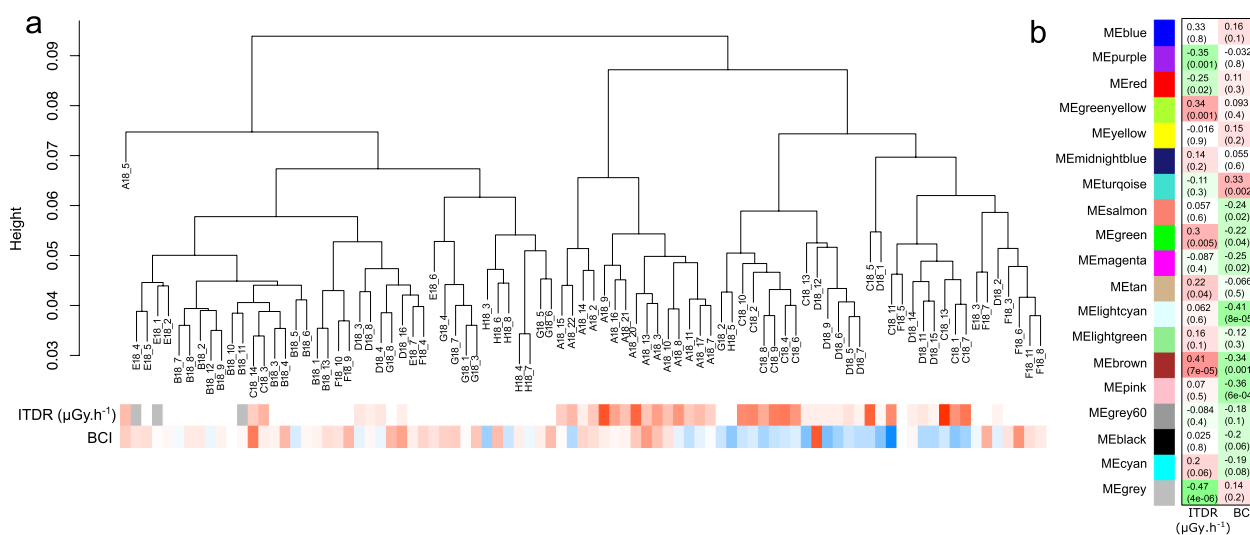


Fig. 6 Association between contig expression and individual variables (BCI and ITDR) by weighted gene co-expression network analysis (WGCNA). **a** Dendrogram of samples based on their Euclidean distance and association with ITDR ($\mu\text{Gy}\cdot\text{h}^{-1}$) and BCI (Ri). The colour codes for the values of the ITDR or BCI: low (blue), moderate (white) and high (red) values. Grey colour indicates missing values. **b** Traits associations with module eigengene correlated to the BCI or ITDR

[56] and bank voles [57]. The fact that population size is smaller in the most contaminated site support the notion that this dynamic is fuelled by a higher mortality or lower reproduction [58]. Thus, migrations and elevated mutations seem not sufficient to offset the genetic diversity loss by bringing new genetic variants in the most contaminated sites of the CEZ.

While we cannot exclude the fact that lower genetic diversity observed in the most contaminated places could be the result of past strong population bottlenecks at the time of the accident (corroborated by the massive functional impacts on amphibians observed at the time of the accident) [23], previous simulations carried out on a mitochondrial marker suggest current small effective population sizes [42]. Current small effective population sizes can be indicative of decreased individual viability in the most contaminated places. Using BCI as an integrative trait to evaluate individual performances linked with energetic status [50], and sometimes proposed as a metric of fitness [59, 60], we showed a negative correlation of this parameter with ITDR. Thus, the exposure to IR is associated with functional modifications in the CEZ. Because of the wide range of mechanisms implicated in BCI variations [61], non-supervised analysis of individual expression variations was performed in order to identify the main functional pathways associated with exposure to IR and mass decrease. The transcriptomic analysis pointed towards an effect of exposure to IR inside the CEZ, as low versus high contaminated sites were discriminated based on gene expression. The gene expression pattern

(observed at the individual level) is explained by ITDR, and this relation remained significant by controlling for geographical distance or genetic distances. The changes in gene expression observed in the CEZ are thus not solely a consequence of genetic differentiation inside the CEZ but rather the result of ITDR differences, despite other environmental characteristics specific to the CEZ cannot be completely excluded. Furthermore, the biological effects detected by the transcriptomic analysis, collectively point toward deregulations of mitochondrial activity and muscle functions in frogs living in the CEZ, which fits with the known effects of chronic exposures to IR [62–65].

The results of the gene expression analysis suggest a substantial role of radioactive contamination on functional differences between individuals, impacting notably major energetic pathways. The proximate mechanism at the origin of these modifications remains unknown. Indeed, it could be both the effect of plastic changes and an adaptive response on a trait selected since several generations, as gene expression levels are known to be heritable and thus potentially under selection [66, 67]. Genetic drift induced a lower range of transmitted variations, which can have functional consequences on which selection can act and thus render the plasticity particularly valuable if the functional changes bring individual closer to a fitness optimum [20]. Moreover, the functional response observed at the transcriptomic level appears constrained (as individuals from the most contaminated places show very similar gene expression patterns) and may thus be indicative

of an adaptive response. A more in-depth analysis of expression diversity in each site for the different contigs could be a way to evaluate the nature of this response and which functions are involved [68, 69]. Although this response seems non-optimal, as it does not compensate the effective population size decline inferred in the most contaminated places, this does not exclude some of these changes from being subject to selection.

Using RNaseq data to analyse SNP represents a promising and cost-effective alternative compared to whole genome sequencing, especially for non-model species for which no reference genome is available. The main limit, however, is that SNP are obtained only from exons that must be expressed at sufficient levels. If we cannot exclude that detection of SNP from gene expression might lead to bias in estimating genetic parameters, recent analysis shows that RNA-seq data can be used with good confidence to detect SNPs [70, 71]. It is not yet possible to discriminate genetic variants only correlated with radiocontamination from loci genuinely under selective process, but many genes involved in response stress and ATPase activity and correlated with ITDR may thus be indicative of selective process on these functions. Several loci strongly associated to ITDR were found to be stop-gained mutations that are predicted to affect the function of several proteins involved in mitochondrial energy production (IDH3G, ALDH2, LDHA). Noteworthy, we observed an increased number of individuals with stop-gained mutations in these genes in the CEZ. Because of the predicted functional deleterious impact of these changes, it is most likely that these genetic variants do not represent cases of selection but rather *de novo* radiation-induced mutations that accumulate in the small population of frogs through inbreeding. Ultimately, these changes are in line with the transcriptomics data that showed changes in ATP metabolic process and glucose metabolism. These changes in metabolism observed at the molecular level in individuals exposed to higher IR doses were correlated to a decreased BCI. In addition, other biological processes were detected by the transcriptomic analysis, including change of expression of genes involved in muscle function and unfolded-protein stress response. As chronic exposure to relative low levels of IR was shown to alter the structure of myofibrils [64], it is possible that our results highlight a disruption of myofibers in frog muscles, an organ with high energy demand and rich in mitochondria. Together, these data collectively point toward a frequent impairment of mitochondrial function, which could cause the degraded body condition of individuals living in the CEZ. Despite congruent functional changes in individuals found in highly contaminated areas of the CEZ, our data do not demonstrate the effects of a selection of traits favourable to

maintain frog population in the CEZ. Rather, the most parsimonious hypothesis is that frog populations inside the CEZ are maintained, in part, by migration from other less contaminated territories.

More than 30 years after the Chernobyl nuclear power plant accident, numerous studies showed contemporary functional modifications in wildlife associated with differences in exposure to IR [35, 72–75]. While generalisation to every exposed species is not possible [31, 76, 77], and the consequences at the ecosystem scale and the mechanisms involved in these changes remain subjected to debate [31], we highlighted the interest of taking an evolutionary perspective to interpret gene expression differences. The functional changes observed in the most contaminated places are expected to influence fitness and thus show the need to study the long-term specific processes that are taking place on wildlife in polluted areas. Mass loss and body condition in particular are expected to influence reproduction by modifying reproduction investment [78, 79]. These fitness costs and the transmission of functional changes through generations should be better addressed in the future, for example with the establishment of field-based reciprocal transplants [80]. We also highlighted the potential key role of energy metabolism in response with radiation exposure from a body of evidence: changes in body condition, differential expression of genes involved in energy metabolism and association of dose rates with genetic variants on genes involved in mitochondria energy production. More generally, we analysed a specific type of response to persistent environmental change which combines an effective small population size coupled to an increase of deleterious mutations in some loci that directed functional plastic changes and viability loss. While these changes might allow the persistence of populations facing the radiocontaminated environment, this study questions the constraints that an eroded genetic diversity places on possible adaptation, and in particular of a ‘genetic assimilation’ which could lead to the fixation of initially plastic phenotypes [81, 82].

Conclusions

The combination of gene expression study, SNP data and population genetic analyses allows us to identify genetic disturbances that were previously inaccessible [42]. In a radio-contaminated environment such as the CEZ, even several decades after the nuclear accident and with total absorbed dose rates that the ICRP has so far estimated to have no deleterious effects on frogs [41, 83], individuals inhabiting the CEZ still display signs of significant disturbances such as genetic erosion, disruption of transcriptomic/metabolic pathways and changes in physiological traits. These changes are consistent with altered

metabolism and fitness costs resulting from long-term radiation exposure, but the causal relationship with radiation remain to be established. Future studies in radio-contaminated areas should use a combination of “omics” data and population genetics analyses to investigate rigorously possible hidden deleterious effects.

Methods

Sampling of *Hyla orientalis* and dosimetry

The individuals used in the present study correspond to sampling made in Ukraine in 2018, as described before [42, 43]. Briefly, *Hyla orientalis* calling male individuals were collected during the night in May–June 2018 from 8 different wetlands in Ukraine, including 6 sites in the CEZ (A18, B18, C18, D18, E18, F18) and two sites 10 km and 45 km east from the Chernobyl NPP (G18 and H18, respectively). Thereafter, these sites are referred to as populations in the sense of “population sample.” The 6 sites inside the CEZ covered a gradient of ambient dose rates ranging from 0.1 to 16.2 $\mu\text{Gy}\cdot\text{h}^{-1}$, while dose rates close to background were measured for the 2 sites outside the CEZ (0.04 $\mu\text{Gy}\cdot\text{h}^{-1}$) (radiometer MKS-AT6130, Atomtex). Absence of trace metallic contamination (including uranium, thorium and plutonium isotopes) in the sites G18 and H18 was confirmed by elemental analysis of water samples (see Elemental Analysis). Individuals were sampled inside the CEZ ($n=73$) and in the 2 non-contaminated sites (G18 and H18) ($n=14$). Three additional individuals were sampled to produce the reference transcriptome (see [Additional Material and Methods](#)). After capture, frogs were kept individually in a wet plastic container overnight, measured, weighted and dissected as described before [42, 43]. The tibia (*gastrocnemius*) muscle was dissected from one leg for all individuals as it is an abundant and easily accessible tissue in frogs, deeply innervated and with high energy demand for locomotion. The other leg was used for reconstruction of the internal contamination (see below). All tibia samples used for RNA extraction were immersed in individual tubes with 1 ml of RNA later (Sigma) and kept in liquid nitrogen. Internal contaminations of ^{90}Sr and ^{137}Cs were measured for each individual in bone and leg muscles, respectively, in the International Radioecology Laboratory (IRL) (Slavutych, Ukraine) to reconstruct the total internal dose, as described before [41, 42]. The individual total dose rate (ITDR) in $\mu\text{Gy}\cdot\text{h}^{-1}$ was reconstructed from both internal and external exposures using EDEN [84], considering both soil and water activities as before. Averaged total dose rates (ATDR) were obtained by computing the mean ITDR per site. All samples were deep frozen until their transport to the Institute for Radiological Protection and Nuclear Safety (IRSN) laboratory (Cadarache, France).

Elemental analysis on water samples

Water samples collected on reference sites (G18 and H18) were acidified to 2% (v/v) HNO_3 with Ultra-Pure JT Baker, 70% (v/v) nitric acid, and analysed for trace metals and major elements with ICP-MS (7800 Agilent, Tokyo, Japan) and ICP-AES (Optima 8300, Perkin Elmer, Waltham, Massachusetts, USA) respectively, according to already published procedures [85] (Additional file 12: Table S10).

RNA extraction and sequencing libraries

The 87 individuals sampled in and outside the CEZ were subjected to RNA-Seq analysis. Total RNA was extracted from tibia muscle with the Quick-DNA/RNA Miniprep kit (Zymoresearch) following manufacturer’s instructions. Integrity and RNA concentrations were assessed on RNA Nano Chips (Bioanalyzer 2100, Agilent). All samples had an RNA integrity index >8 and were used for library preparation using the Stranded mRNA Prep Ligation kit with ITD UD index (Illumina) following instructions. After quality check (size and concentration) on DNA1000 Chips (Bioanalyzer 2100, Agilent), libraries were sequenced on a NovaSeq6000 platform to produce 100 or 150 bases long paired-end reads using S4 flow cells (Clinical Research Sequencing Platform, Broad Institute, MIT, United States). All fastq files were quality checked with an in-house pipeline relying on fastqc and fastx-toolkit.

Transcriptomics analysis

The de novo assembly of *Hyla orientalis* transcriptome was made with Trinity [86] using the complete dataset obtained from five different tissue type (tibia muscle, heart, eye, brain and testis) obtained from 3 additional individuals sampled in non-contaminated site. Completeness of *Hyla orientalis* transcriptome assembly was assessed by BUSCO [51] analysis. We searched similarities to SwissProt [87] and UniRef90 [88] ($E < 1e^{-3}$) using blastx and blastp [89] with the assembled contigs and predicted ORF respectively. All annotations were collected in a mySQLite relational database using *Trinotate* [90] for annotating putative functional characteristics. The quality of the assembled contigs was confirmed by comparing the sequence of the predicted ploypeptides to proteomics data obtained by Mass spectrometry from 3 muscles of the same individuals. Proteins were extracted as before [64]. Peptides were analysed using liquid nanochromatography online in front of an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) as previously describe [91]. Transcript abundance estimation were mapped against the de novo assembled

Hyla orientalis transcriptome using Bowtie2 [92] and quantified by RSEM [93]. Differential analysis was made by Deseq2 [94]. Dose–response modelling and BMR (benchmark response) estimation were computed with DRomics [54]. Weighted gene co-expression network analysis was performed with WGCNA [55]. See Additional file 3: Supplementary Material and Methods and Additional file 2: Fig. S1 for details on methods for the transcriptomics data analysis.

Variant detection

Uniquely mapped reads were processed for variant calling as described in GATK’s best practices for RNA-Seq data [95, 96]. Briefly, mate information was fixed, duplicates marked and reads split with the GATK function SplitNCGarReads. No base recalibration and base quality reassignment could be performed as prior SNPs information are missing for this species. Germline variants were detected in each sample using GATK HaplotypeCaller 4.2.0.0 with the options `-native-pair-hmm-threads 46 -pcr-indel-model NONE -ERC GVCF`. Individual vcf files were then used for joint genotyping with the GATK’s functions GenomicsDBImport and GenotypeGVCFs to produce a single gvcf file. The following filtering steps were then performed: (i) vcfutils [97] was used with the options `-remove-indels -max-missing 0.75 -recode` to remove sites were at most 25% of the genotypes were missing, (ii) we then applied the following filters recommended by GATK’s best practices for RNA-Seq SNPs calling: `-filter “QD<2.0” -filter-name “QD2” -filter “QUAL<30.0” -filter-name “QUAL30” -filter “SOR>3.0” -filter-name “SOR3” -filter “FS>60.0” -filter-name “FS60” -filter “FS>30.0” -filter-name “FS30” -filter “MQ<40.0” -filter-name “MQ40” -filter “MQRankSum<-12.5” -filter-name “MQRankSum-12.5” -filter “ReadPosRankSum<-8.0” -filter-name “ReadPosRankSum-8”`.

Population genetics summary statistics

Nucleotide diversity (π) for each population is the mean of nucleotide diversity for each contig calculated as the average number of nucleotide differences ($\theta\pi$) divided by the total number of positions (segregating or not segregating sites) except positions with missing data in any individual. The individual heterozygosity ratio was calculated as the number of heterozygous positions on the number of segregating sites used for an individual. π and the individual heterozygosity were computed with DnaSP [98]. Expected (H_e) and observed (H_o) Heterozygosity of each population for each contig were computed with PLINK [99] and inbreeding coefficient (Fis) as $1 - (H_o/H_e)$ [100]. We then analysed the distribution of Fis values for the different loci for each population. To evaluate the

contribution of loci with extreme values, we compared Fis between populations solely for loci at the Hardy–Weinberg equilibrium.

Population structure

Genetic variability was computed by pairwise Hedrick $G'st$ [101] analysis between all populations with vcfR v.1.12.0 [102]. Multivariate analyses were performed among bi-allelic SNPs to take into account the different loci. Principal component analysis (PCA) and discriminant analysis of principal component (DAPC) were conducted with the adegenet v.2.1.3 package [103]. For DAPC, eight principal components were retained representing 21% of the total variance to keep a sufficient power of discrimination between observed and random effects and to avoid overfitting (a-score criterion). We retained and interpreted the two first discriminant functions which were the most informative. Group memberships of DAPC were represented with a “STRUCTURE-like” barplot to investigate differentiation between populations and their level of admixture. In addition, ancestry estimation was obtained from the individual SNP data with the algorithm *admixture* using $k=7$ [104]. In order to make inferences about the potential effective size differences between populations, we estimated individual pairwise kinship coefficient [105] for individuals which are affiliated to more than 95% to their initial population (to get rid of potentially recent migrants). The identification of genetic clusters without a priori was made with k-means clustering with all PCs and a number of cluster $k=8$, corresponding to the eight populations sampled as a maximum of cluster.

Identification of loci potentially under selection

To identify loci potentially under selection along the contamination gradient, we used a Latent Factor Mixed Model (LFMM) implemented in LEA v.2.8.0 package [106]. This model provides a genome-wide ecological association study, by fitting a statistical model including population structure, which permits to test the association between genetic polymorphism and an environmental variable. Here, by running first non-negative matrix factorisation algorithms (sNMF), 3 clusters were selected to optimise the cross-entropy criterion. LFMM were then performed with 5 runs (iterations=6000, burnin=3000) with the ITDR as environmental variable and outliers identified using q -values (FDR, Benjamini-Hochberg) < 0.05 . Genomic variant annotations and functional effect prediction were made with SnpEff [107]. Biallelic SNP outliers predicted to cause homozygous stop-gained mutations were tested for deviating from the HW equilibrium using the R package HardyWeinberg version 1.4.1 using the HWExact test option.

Statistical analysis of the link between ionizing radiation and body condition, genetic diversity, and expression

Body condition of Eastern tree frogs was estimated with the residuals of the regression of $\log_{10}(\text{body mass})$ against $\log_{10}(\text{snout-vent length})$ (Ri) [108, 109]. Linear mixed models were used to test the link between $\log(\text{ITDR})$ and body condition index with site as random factor. Linear models were used to test the link between $\log(\text{ITDR})$ and individual heterozygosity. Correlations between genetic diversity indices (π and F_{is}) and ATDR were tested with non-parametric Spearman's rank tests. To investigate potential changes in variance in relatedness along the contamination gradient, correlation between ATDR and the residuals of relatedness for each site, calculated as the absolute value of relatedness from median of relatedness of each site, was tested with non-parametric Kendall's test. At the population level, correlation between $\log(\text{geographical distances})$ (in meter, m) and genetic distances ($G'st/(1-G'st)$) were tested using Mantel's test [110]. The dependence of this correlation on pairwise ATDR absolute difference was tested using partial-Mantel test. At the individual level, Mantel tests were used to test the correlation between cophenetic distances (i.e. interindividual similarity) obtained by the hierarchical clustering of 5735 differentially expressed contig and $\log(\text{ITDR})$ or $\log(\text{geographical distances})$ (see also [Additional Material and Methods](#)). The dependence of these correlations on $\log(\text{geographical distances})$ and genetic distances respectively was tested with partial Mantel tests. Mantel and partial-Mantel tests were carried out on *vegan* v.2.5–6 package [111] and significance estimated by permutation's test using 9999 permutations. Threshold of significance were all set to $p < 0.05$. All tests were carried out on R version 3.6.1 (A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>).

Abbreviations

ATDR	Average total dose rate
BCI	Body condition index
CEZ	Chernobyl Exclusion Zone
DAPC	Discriminant analysis of principal component
FDR	False discovery rate
Fis	Inbreeding coefficient
Gst	Coefficient of genetic differentiation
HW	Hardy-Weinberg
ICRP	International Commission on Radiological Protection
ITDR	Individual total dose rate
IR	Ionizing radiation
LFMM	Latent Factor Mixed Model
PCA	Principal component analysis
RNASeq	RNA high-throughput sequencing
RNA	Ribonucleic acid
SNP	Single-nucleotide polymorphism
WGCNA	Weighted correlation network analysis
π	Nucleotide diversity

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-023-01659-2>.

Additional file 1: Table S1. Sampling sites, doses and individual parameters measured in 87 sampled frogs including weight, length, width, BCI (Ri), Individual Total Dose Rate (ITDR) and Average Total Dose Rate (ATDR).

Additional file 2: Fig. S1. Experimental design of the transcriptomic analysis on *Hyla orientalis*. **Fig. S2.** Contigs expression and population structure analysis. a. PCA on RNAseq expression data. b. DAPC on genetic variants obtained from the 87 individuals. c. Results of the K-mean clustering of individuals based on genetic polymorphisms. d. Admixture analysis made on the SNP from the 87 individuals to assess ancestry coefficient.

Fig. S3. Bootstrap analysis after hierarchical clustering of 5,735 contigs expressed in 87 individuals. AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) value are represented at each node in red and green respectively. AU p-value (computed by multiscale bootstrap resampling) is a good approximation of unbiased p-value. Branches with confidence intervals > 90% are highlighted in red. **Fig. S4.** Comparison of the differential expression analysis ($n = 17,323$ contigs) using G18, H18 or F18 populations as references. The correlation of the FC obtained by using the different reference sites is indicated in each comparative analysis. The Pearson correlation coefficient ρ is indicated at the top. **Fig. S5.** a. Unsupervised hierarchical clustering of normalised expression of 5,735 contigs differentially expressed in at least one population (using G18 individuals as control) after removing the effects of genetic distances as covariables. The expression levels of contigs (as Row Z-score) are represented as follows: low expression (blue), moderate expression (white) and high expression (red). b. Enrichments of Gene Ontology Biological Process in the different clusters obtained by hierarchical clustering. The numbers in brackets correspond to the total number of contigs in each cluster (bottom, x-axis).

Fig. S6. Association between contig expression and individual variables (BCI and ITDR) by WGCNA after adjusting the expression for genetic distances and removing low abundance genes ($n = 13,641$). Traits associations with module eigengene correlated to the BCI and ITDR are indicated for each module (correlation coefficient and p-value are indicated).

Additional file 3. Supplementary material and methods.

Additional file 4: Table S2. NGS quality control of RNAseq 87 individuals, including RNA index Number (RIN), read coverage and read quality (Phred score as Q30).

Additional file 5: Table S3. Summary of SNPs correlated with ITDR in the CEZ. The result of tblastx giving the closest human Uniprot orthologue is indicated, as well as the results from SNP calling with GATK, variant prediction from SnpEff and the DEG obtained by DESeq2.

Additional file 6: Table S4. a. List of the 159 SNPs in the 143 different contigs presenting a stop gained variants for the three level of contamination: Red (populations in the centre of the CEZ exposed to "higher" dose rate: A,C,D), Yellow ("medium" dose rate for populations BEF), Green ("lower" dose rates, for populations G,H). We indicate the p-value of the Fisher exact test for hardy-Weinberg equilibrium: green circle not significant, yellow circle significant with p-value ranging from 0.01 to 0.05, red circle significant with p-value < 0.01. b. Number of significant Fisher exact test p-value for the three different radioactive contamination area (high = red; yellow = medium; green = low). c. Genotype distribution (Wt = wild type; + stop gained mutation) for the different genes presenting a p-value < 0.01 and for the three different radioactive contamination area.

Additional file 7: Table S5. Functional enrichment of differentially expressed contigs in common after categorisation of population living in highly contaminated sites (named as cluster 1: A18, C18, D18 versus G18) and low contaminated sites (named as cluster 2: B18, E18, F18 versus G18). a. KEGG, b. Molecular Function, c. Biological Process.

Additional file 8: Table S6. Functional enrichment in each of the 9 clusters obtained by hierarchical clustering of differentially expressed contigs ($n = 5,735$). a. KEGG, b. Biological process, c. Molecular Function.

Additional file 9: Table S7. Enrichment of GO Biological Process after dose–response modelling of contig expression by Dromics.

Additional file 10: Table S8. Enrichment of GO Biological Process and GO Molecular functions in module eigengenes obtained from weighted gene co-expression network analysis.

Additional file 11: Table S9. Enrichment of GO Biological Process functions in module eigengenes obtained from weighted gene co-expression network analysis using the matrix of expression filtered for lowly expressed genes and corrected by genetic distances.

Additional file 12: Table S10. Concentration of heavy metals and radionuclides measured on the two reference sites near Slavutykh (G18 and H18). Recommended guidelines are indicated for each element (when existing). Na: Not available.

Acknowledgements

We are especially thankful to Dr. Sergey Gaschack from the Chernobyl Center for Nuclear Safety in Ukraine who organised the logistic, advised and participated in the capture of tree frogs and conducted radioisotope analyses. We want to thank Loïc Carasco (IRSN/SRTE/LR2T) for its help in elemental analysis as well as Adrien Moulin and Philippe Beudeart for their help in grid computing.

Authors' contributions

OA, AG and JMB designed the study. CC conducted the experiments and performed analysis of the BCI and genetics data. OA performed the analysis of the transcriptomics data and produced the SNP data. MLDM performed the dose–response analysis. GO, PB and JMB performed the field work. CC and JMB performed the reconstruction of the internal and external dose. LC, EB, SA and SF prepared the samples for proteomics and analysed the data. JA advised on the proteomics data analysis on environmental sentinel organisms. EG, IG and AT performed the regression on expression matrix to remove the effects of genetic variabilities and advised on the statistical analysis for the whole manuscript. All authors contributed in writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the research framework of IRSN, CNRS NEEDS-Environment through the project Tchernob'Hyla (OA), ECCOREV (Hyla Project) (OA) and EC2CO (JMB, SF, OA), the Swedish Radiation Protection Agency (SSM2018-2038) (GO), the Spanish Ministry of Science and innovation (RYC-2016–20656) (GO), Carl Tryggers Foundation (16:344) (GO, PB), and Marie Skłodowska-Curie (METAGE 797879) (PB).

Availability of data and materials

All data generated or analysed during this study are included in this published article, its supplementary information files and publicly available repositories. RNASeq data are available under the Gene Expression Omnibus (GEO) accession number GSE211060 at the address: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211060>. [112].

Declarations

Ethics approval and consent to participate

Animals were collected under the permit of the Ministry of Ecology and Natural Resources of Ukraine No. 517, 21042016. All methods and experimental protocols were conducted in accordance with ethical guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institut de Radioprotection Et de Sûreté Nucléaire (IRSN), PSE-ENV/SRTE/LECO, Cadarache, France. ²PSE-SANTE/SESANE/LRTox, Fontenay Aux Roses,

France. ³UMR 1467 RECOVER, Aix-Marseille Université, INRAE, Centre Saint-Charles, Marseille, France. ⁴Laboratoire Des Signaux Et Systèmes, Université Paris-Saclay, CNRS, CentraleSupélec, 91190 Gif-Sur-Yvette, France. ⁵Laboratoire de Biométrie Et Biologie Evolutive, UMR 5558, Université de Lyon, Université Lyon 1, CNRS, VetAgro Sup, Villeurbanne, France. ⁶Aix-Marseille University, Inserm, CNRS, Institut Paoli-Calmettes, CRCM, Marseille Proteomics, Marseille, France. ⁷Animal Ecology, Department of Ecology and Genetics, Evolutionary Centre, Uppsala University, 75236 Uppsala, Sweden. ⁸Doñana Biological Station (CSIC), Seville, Spain. ⁹IMIB-Biodiversity Research Institute, University of Oviedo, 33600 Mieres-Asturias, Spain. ¹⁰Zoology Unit, Department of Biology of Organisms and Systems, University of Oviedo, 33071 Oviedo-Asturias, Spain. ¹¹Département Médicaments Et Technologies Pour La Santé (DMTS), Université Paris-Saclay, CEA, INRAE, SPI, Bagnols-Sur-Cèze, France.

Received: 16 January 2023 Accepted: 3 July 2023

Published online: 31 July 2023

References

- Pacifici M, Foden WB, Visconti P, Watson JEM, Butchart SHM, Kovacs KM, et al. Assessing species vulnerability to climate change. *Nature Clim Change*. 2015;5:215–24.
- Farnsworth KD, Albantakis L, Caruso T. Unifying concepts of biological function from molecules to ecosystems. *Oikos*. 2017;126:1367–76.
- Kelly M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philos Trans R Soc Lond B Biol Sci*. 2019;374:20180176.
- Fox RJ, Donelson JM, Schunter C, Ravasi T, Gaitán-Espitia JD. Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philos Trans R Soc Lond B Biol Sci*. 2019;374:20180174.
- Diamond SE, Martin RA. The interplay between plasticity and evolution in response to human-induced environmental change. *F1000Res*. 2016;5:2835.
- Stockwell CA, Hendry AP, Kinnison MT. Contemporary evolution meets conservation biology. *Trends Ecol Evol*. 2003;18:94–101.
- Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. Conservation genetics in transition to conservation genomics. *Trends Gen*. 2010;26:177–87.
- Lambert MR, Brans KI, Roches SD, Donihue CM, Diamond SE. Adaptive evolution in cities: progress and misconceptions. *Trends Ecol Evol*. 2021;36:239–57.
- Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl*. 2014;7:1–14.
- Loria A, Cristescu ME, Gonzalez A. Mixed evidence for adaptation to environmental pollution. *Evol Appl*. 2019;12:1259–73.
- Crispo E, Moore J-S, Lee-Yaw JA, Gray SM, Haller BC. Broken barriers: human-induced changes to gene flow and introgression in animals. *BioEssays*. 2011;33:508–18.
- Jha S. Contemporary human-altered landscapes and oceanic barriers reduce bumble bee gene flow. *Mol Ecol*. 2015;24:993–1006.
- Bijlsma R, Loeschcke V. Genetic erosion impedes adaptive responses to stressful environments. *Evol Appl*. 2012;5:117–29.
- Jump AS, Marchant R, Peñuelas J. Environmental change and the option value of genetic diversity. *Trends Plant Sci*. 2009;14:51–8.
- Eeva T, Belskii E, Kuranov B. Environmental pollution affects genetic diversity in wild bird populations. *Mutat Res*. 2006;608:8–15.
- Halligan DL, Keightley PD. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Syst*. 2009;40:151–72.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics*. 1998;148:1667–86.
- Mehlhoff JD, Stearns FW, Rohm D, Wang B, Tsou E-Y, Dutta N, et al. Collateral fitness effects of mutations. *Proc Natl Acad Sci U S A*. 2020;117:11597–607.
- Lynch M, Conery J, Bürger R. Mutational meltdowns in sexual populations. *Evolution*. 1995;49:1067–80.
- Ghalambor CK, McKay JK, Carroll SP, Reznick DN. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol*. 2007;21:394–407.

21. Coutellec M-A, Barata C. An introduction to evolutionary processes in ecotoxicology. *Ecotoxicology*. 2011;20:493–6.
22. Bickham JW. The four cornerstones of Evolutionary Toxicology. *Ecotoxicology*. 2011;20:497–502.
23. Geras'kin SA, Fesenko SV, Alexakhin RM. Effects of non-human species irradiation after the Chernobyl NPP accident. *Environ Int*. 2008;34:880–97.
24. Geras'kin SA, Fesenko SV, Volkova PY, Isamov NN. What have we learned about the biological effects of radiation from the 35 years of analysis of the consequences of the Chernobyl NPP accident? *Biol Bull Russ Acad Sci*. 2021;48:2105–26.
25. Møller AP, Mousseau TA. Biological consequences of Chernobyl: 20 years on. *Trends Ecol Evol*. 2006;21:200–7.
26. Smith JT, Beresford NA. Chernobyl - catastrophe and consequences. In: Smith JT, Beresford NA, editors. *Chernobyl — Catastrophe and Consequences*. Berlin: Springer; 2005. p. 289–305. Available from: https://doi.org/10.1007/3-540-28079-0_9. Cited 4 Nov 2022.
27. Alexis-Martin B, Davies T. Towards nuclear geography: zones, bodies, and communities. *Geogr Compass*. 2017;11:e12325.
28. Smith JT, Beresford NA. Chernobyl - catastrophe and consequences. In: Smith JT, Beresford NA, editors. *Chernobyl — catastrophe and consequences*. Berlin: Springer; 2005. p. 1–34. Available from: https://doi.org/10.1007/3-540-28079-0_1. Cited 4 Nov 2022.
29. Deryabina TG, Kuchmel SV, Nagorskaya LL, Hinton TG, Beasley JC, Lerebours A, et al. Long-term census data reveal abundant wildlife populations at Chernobyl. *Curr Biol*. 2015;25:R824–826.
30. Bréchinac F, Oughton D, Mays C, Barnthouse L, Beasley JC, Bonisoli-Alquati A, et al. Addressing ecological effects of radiation on populations and ecosystems to improve protection of the environment against radiation: agreed statements from a Consensus Symposium. *J Environ Radioact*. 2016;158–159:21–9.
31. Beresford NA, Horemans N, Copplestone D, Raines KE, Orizaola G, Wood MD, et al. Towards solving a scientific controversy - the effects of ionising radiation on the environment. *J Environ Radioact*. 2020;211:106033.
32. Boratyński Z, Lehmann P, Mappes T, Mousseau TA, Møller AP. Increased radiation from Chernobyl decreases the expression of red colouration in natural populations of bank voles (*Myodes glareolus*). *Sci Rep*. 2014;4:7141.
33. Lehmann P, Boratyński Z, Mappes T, Mousseau TA, Møller AP. Fitness costs of increased cataract frequency and cumulative radiation dose in natural mammalian populations from Chernobyl. *Sci Rep*. 2016;6:19974.
34. Kivisaari K, Boratyński Z, Lavrinienko A, Kesäniemi J, Lehmann P, Mappes T. The effect of chronic low-dose environmental radiation on organ mass of bank voles in the Chernobyl Exclusion Zone. *Int J Radiat Biol*. 2020;96:1254–62.
35. Jernfors T, Kesäniemi J, Lavrinienko A, Mappes T, Milinevsky G, Møller AP, et al. Transcriptional upregulation of DNA damage response genes in bank voles (*Myodes glareolus*) inhabiting the Chernobyl Exclusion Zone. *Front Environ Sci*. 2018;5:95. Available from: <https://www.frontiersin.org/article/10.3389/fenvs.2017.00095>. Cited 3 Feb 2022.
36. Mappes T, Boratyński Z, Kivisaari K, Lavrinienko A, Milinevsky G, Mousseau TA, et al. Ecological mechanisms can modify radiation effects in a key forest mammal of Chernobyl. *Ecosphere*. 2019;10:e02667.
37. Vos CC, Braak CJFT, Nieuwenhuizen W. Incidence function modelling and conservation of the tree frog *Hyla arborea* in the Netherlands. *Ecological Bulletins*. 2000;165–80.
38. Ssc) ISASG (IUCN). IUCN Red List of Threatened Species: *Hyla orientalis*. IUCN Red List of Threatened Species. 2021. Available from: <https://www.iucnredlist.org/en>. Cited 4 Nov 2022.
39. Maselli V, Polese G, Rippa D, Ligrone R, Kumar Rastogi R, Fulgione D. Frogs, sentinels of DNA damage induced by pollution in Naples and the neighbouring Provinces. *Ecotoxicol Environ Saf*. 2010;73:1525–9.
40. Othman MS, Khonsue W, Kitana J, Thirakhupt K, Robson MG, Kitana N. Cadmium accumulation in two populations of rice frogs (*Fejervarya limnocharis*) naturally exposed to different environmental cadmium levels. *Bull Environ Contam Toxicol*. 2009;83:703–7.
41. Burraco P, Car C, Bonzom J-M, Orizaola G. Assessment of exposure to ionizing radiation in Chernobyl tree frogs (*Hyla orientalis*). *Sci Rep*. 2021;11:20509.
42. Car C, Gilles A, Armant O, Burraco P, Beaugelin-Seiller K, Gashchak S, et al. Unusual evolution of tree frog populations in the Chernobyl Exclusion Zone. *Evol Appl*. 2022;15:203–19.
43. Burraco P, Bonzom J-M, Car C, Beaugelin-Seiller K, Gashchak S, Orizaola G. Lack of impact of radiation on blood physiology biomarkers of Chernobyl tree frogs. *Front Zool*. 2021;18:33.
44. Schirmer K, Fischer BB, Madureira DJ, Pillai S. Transcriptomics in ecotoxicology. *Anal Bioanal Chem*. 2010;397:917–23.
45. Whitehead A, Triant DA, Champlin D, Nacci D. Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Mol Ecol*. 2010;19:5186–203.
46. Banerjee SM, Stoll JA, Allen CD, Lynch JM, Harris HS, Kenyon L, et al. Species and population specific gene expression in blood transcriptomes of marine turtles. *BMC Genomics*. 2021;22:346.
47. Depardieu C, Gérardi S, Nadeau S, Parent GJ, Mackay J, Lenz P, et al. Connecting tree-ring phenotypes, genetic associations and transcriptomics to decipher the genomic architecture of drought adaptation in a widespread conifer. *Mol Ecol*. 2021;30:3898–917.
48. Rougeux C, Gagnaire P-A, Praebel K, Seehausen O, Bernatchez L. Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Mol Ecol*. 2019;28:4388–403.
49. Albecker MA, Stuckert AMM, Balakrishnan CN, McCoy MW. Molecular mechanisms of local adaptation for salt-tolerance in a treefrog. *Mol Ecol*. 2021;30:2065–86.
50. Brodeur JC, Damonte MJ, Vera Candioti J, Poliserpi MB, D'Andrea MF, Bahl MF. Frog body condition: basic assumptions, comparison of methods and characterization of natural variability with field data from *Leptodactylus latrans*. *Ecol Ind*. 2020;112:106098.
51. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
52. Glancy B, Balaban RS. Energy metabolism design of the striated muscle cell. *Physiol Rev*. 2021;101:1561–607.
53. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet*. 2013;93:641–51.
54. Larras F, Billoir E, Baillard V, Siberchicot A, Scholz S, Wubet T, et al. DRomics: a Turnkey Tool to support the use of the dose-response framework for omics data in ecological risk assessment. *Environ Sci Technol*. 2018;52:14461–8.
55. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
56. Møller AP, Hobson KA, Mousseau TA, Peklo AM. Chernobyl as a population sink for barn swallows: tracking dispersal using stable-isotope profiles. *Ecol Appl*. 2006;16:1696–705.
57. Kesäniemi J, Boratyński Z, Danforth J, Itam P, Jernfors T, Lavrinienko A, et al. Analysis of heteroplasmy in bank voles inhabiting the Chernobyl Exclusion Zone: a commentary on Baker et al. (2017) "Elevated mitochondrial genome variation after 50 generations of radiation exposure in a wild rodent." *Evol Appl*. 2018;11(820):6.
58. Dias PC. Sources and sinks in population biology. *Trends Ecol Evol*. 1996;11:326–30.
59. Janin A, Léna J-P, Joly P. Beyond occurrence: body condition and stress hormone as integrative indicators of habitat availability and fragmentation in the common toad. *Biol Cons*. 2011;144:1008–16.
60. Cogălniceanu D, Stănescu F, Székely D, Topliceanu T-S, Iosif R, Székely P. Age, size and body condition do not equally reflect population response to habitat change in the common spadefoot toad *Pelobates fuscus*. *PeerJ*. 2021;9:e11678.
61. Wilder SM, Raubenheimer D, Simpson SJ. Moving beyond body condition indices as an estimate of fitness in ecological and evolutionary studies. *Funct Ecol*. 2016;30:108–15.
62. Kam WWY, Banati RB. Effects of ionizing radiation on mitochondria. *Free Radic Biol Med*. 2013;65:607–19.
63. Kostyuk SV, Proskurnina EV, Konkova MS, Abramova MS, Kalianov AA, Ershova ES, et al. Effect of low-dose ionizing radiation on the expression of mitochondria-related genes in human mesenchymal stem cells. *Int J Mol Sci*. 2021;23:261.
64. Murat El Houdigui S, Adam-Guillermin C, Loro G, Arcanjo C, Frelon S, Floriani M, et al. A systems biology approach reveals neuronal and muscle developmental defects after chronic exposure to ionising radiation in zebrafish. *Sci Rep*. 2019;9:20241.
65. Murat El Houdigui S, Adam-Guillermin C, Armant O. Ionising radiation induces promoter DNA hypomethylation and perturbs transcriptional

- activity of genes involved in morphogenesis during gastrulation in zebrafish. *Int J Mol Sci.* 2020;21:E4014.
66. Whitehead A, Crawford DL. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol.* 2006;15:1197–211.
 67. Leder EH, McCairns RJS, Leinonen T, Cano JM, Viitaniemi HM, Nikinmaa M, et al. The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. *Mol Biol Evol.* 2015;32:674–89.
 68. Yan J, Song Z, Xu Q, Kang L, Zhu C, Xing S, et al. Population transcriptomic characterization of the genetic and expression variation of a candidate progenitor of *Miscanthus* energy crops. *Mol Ecol.* 2017;26:5911–22.
 69. Herrmann M, Ravindran SP, Schwenk K, Cordellier M. Population transcriptomics in *Daphnia*: the role of thermal selection. *Mol Ecol.* 2018;27:387–402.
 70. Rogier O, Chateigner A, Amanzougarene S, Lesage-Descauses M-C, Balzergue S, Brunaud V, et al. Accuracy of RNAseq based SNP discovery and genotyping in *Populus nigra*. *BMC Genomics.* 2018;19:909.
 71. Zhao Y, Wang K, Wang W, Yin T, Dong W, Xu C. A high-throughput SNP discovery strategy for RNA-seq data. *BMC Genomics.* 2019;20:160.
 72. Kesäniemi J, Jernfors T, Lavrinienko A, Kivisaari K, Kiljunen M, Mappes T, et al. Exposure to environmental radionuclides is associated with altered metabolic and immunity pathways in a wild rodent. *Mol Ecol.* 2019;28:4620–35.
 73. Mustonen V, Kesäniemi J, Lavrinienko A, Tukalenko E, Mappes T, Watts PC, et al. Fibroblasts from bank voles inhabiting Chernobyl have increased resistance against oxidative and DNA stresses. *BMC Cell Biol.* 2018;19:17.
 74. Lerebours A, Gudkov D, Nagorskaya L, Kaglyan A, Rizewski V, Leshchenko A, et al. Impact of environmental radiation on the health and reproductive status of fish from Chernobyl. *Environ Sci Technol.* 2018;52:9442–50.
 75. Arnaise S, Shykoff JA, Møller AP, Mousseau TA, Giraud T. Anther-smut fungi from more contaminated sites in Chernobyl show lower infection ability and lower viability following experimental irradiation. *Ecol Evol.* 2020;10:6409–20.
 76. Fuller N, Ford AT, Nagorskaya LL, Gudkov DI, Smith JT. Reproduction in the freshwater crustacean *Asellus aquaticus* along a gradient of radionuclide contamination at Chernobyl. *Sci Total Environ.* 2018;628–629:11–7.
 77. Fuller N, Smith JT, Nagorskaya LL, Gudkov DI, Ford AT. Does Chernobyl-derived radiation impact the developmental stability of *Asellus aquaticus* 30 years on? *Sci Total Environ.* 2017;576:242–50.
 78. Brandt LSE, Greenfield MD. Condition-dependent traits and the capture of genetic variance in male advertisement song. *J Evol Biol.* 2004;17:821–8.
 79. Roznik EA, Sapsford SJ, Pike DA, Schwarzkopf L, Alford RA. Condition-dependent reproductive effort in frogs infected by a widespread pathogen. *Proc Biol Sci.* 2015;282:20150694.
 80. Forgiione ME, Brady SP. Road salt is more toxic to wood frog embryos from polluted ponds. *Environ Pollut.* 2022;296:118757.
 81. Pigliucci M, Murren CJ, Schlichting CD. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol.* 2006;209:2362–7.
 82. Waddington CH. Genetic Assimilation. In: Caspari EW, Thoday JM, editors. *Advances in Genetics.* Academic Press. 196. p. 257–93. Available from: <https://www.sciencedirect.com/science/article/pii/S0065266008601194>. Cited 18 Feb 2022.
 83. ICRP publication 108. Environmental Protection - the Concept and Use of Reference Animals and Plants. 2008. p. 38. Available from: <https://www.icrp.org/publication.asp?id=icrp%20publication%20108>.
 84. Beaugelin-Seiller K, Jasserand F, Garnier-Laplace J, Gariel JC. Modeling radiological dose in non-human species: principles, computerization, and application. *Health Phys.* 2006;90:485–93.
 85. Mounicou S, Frelon S, Le Guernic A, Eb-Levadoux Y, Camilleri V, Février L, et al. Use of fish otoliths as a temporal biomarker of field uranium exposure. *Sci Total Environ.* 2019;690:511–21.
 86. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 2011;29:644–52.
 87. Boeckmann B, Blatter M-C, Famiglietti L, Hinz U, Lane L, Roehert B, et al. Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *C R Biol.* 2005;328:882–99.
 88. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, the UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics.* 2015;31:926–32.
 89. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
 90. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 2017;18:762–76.
 91. Dubois C, Pophillat M, Audebert S, Fourquet P, Lecomte C, Dubourg N, et al. Differential modification of the *C. elegans* proteome in response to acute and chronic gamma radiation: Link with reproduction decline. *Sci Total Environ.* 2019;676:767–81.
 92. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
 93. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
 94. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
 95. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytisky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
 96. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
 97. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
 98. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34:3299–302.
 99. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
 100. Nei M, Chesser RK. Estimation of fixation indices and gene diversities. *Ann Hum Genet.* 1983;47:253–9.
 101. Hedrick PW. A standardized genetic differentiation measure. *Evolution.* 2005;59:1633–8.
 102. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour.* 2017;17:44–53.
 103. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008;24:1403–5.
 104. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011;12:246.
 105. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26:2867–73.
 106. Frichot E, François O. LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol.* 2015;6:925–9.
 107. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 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).* 2012;6:80–92.
 108. Green AJ. Mass/length residuals: measures of body condition or generators of spurious results? *Ecology.* 2001;82:1473–83.
 109. Plăiașu R, Hartel T, Băncilă RI, Cogălniceanu D, Smets J. Comparing three body condition indices in amphibians: a case study of yellow-bellied toad *Bombina variegata*. *Amphibia-Reptilia.* 2010;31:558–62.
 110. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967;27:209–20.
 111. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
 112. Armant O, Car C, Frelon S, Camoin L. Population transcriptogenomics highlights impaired metabolism and small population sizes in tree frogs living in the Chernobyl Exclusion Zone. *Gene Expression Omnibus.* 2023. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211060>. Cited 15 Jun 2023.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.