bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enhanced plasticity of programmed DNA elimination boosts adaptive potential in suboptimal environments

Valerio Vitali, Rebecca Hagen, and Francesco Catania*

Institute for Evolution and Biodiversity, University of Münster, Hüfferstrasse 1, 48149 Münster, Germany

* To whom correspondence should be addressed
Francesco Catania,
Hüfferstrasse 1,
48149 Münster, Germany
Phone: +49-251-8321222
francesco.catania@uni-muenster.de

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The impact of ecological changes on the development of new somatic genomes has thus far been neglected. This oversight yields an incomplete understanding of the mechanisms that underlie environmental adaptation and can be tackled leveraging the biological properties of ciliates. When *Paramecium* reproduces sexually, its polyploid somatic genome regenerates from the germline genome via a developmental process, Programmed DNA elimination (PDE), that involves the removal of thousands of ORF-interrupting germline sequences. Here, we demonstrate that exposure to sub-optimal temperatures impacts PDE efficiency, prompting the emergence of hundreds of alternative DNA splicing variants that dually embody cryptic (germline) variation and *de novo* induced (somatic) mutations. In contrast to trivial biological errors, many of these alternative DNA isoforms display a patterned genomic topography, are epigenetically controlled, inherited trans-somatically, and under purifying selection. Developmental thermoplasticity in *Paramecium* is a likely source of evolutionary innovation.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Developmental plasticity—the environmentally induced phenotypic variance associated with alternative developmental trajectories—has been proposed to fuel adaptive evolution by initiating phenotypic changes **(West-Eberhard 2005; Uller et al. 2018)**. Exploring the molecular mechanisms that underlie developmental plasticity can reveal a direct link between environmental changes and phenotypic differentiation, shedding light on how variation can surface from a single genotype in a stressful environment. This knowledge has important consequences for current understanding of evolutionary processes and human health **(Lea et al. 2017b; Lea et al. 2017a).**

Previous studies in flies, plants, fungi, and vertebrates suggest that environmental changes that alter the molecular chaperone Hsp90's buffering capacity during development can unlock cryptic genetic variation and boost phenotypic diversification **(Rutherford and Lindquist 1998; Queitsch et al. 2002; Yeyati et al. 2007; Jarosz and Lindquist 2010; Rohner et al. 2013)**. These observations substantiate an evolutionary model where cryptic developmental variation, which is revealed in response to environmental stress, might become genetically assimilated **(Waddington 1953)**. An alternative mechanism that links genetic and phenotypic variation via environmental stress has also been proposed. Recent studies in flies suggest that environmental stress, rather than exposing cryptic variation, may induce *de novo* mutations, DNA deletions and transposon insertions **(Fanti et al. 2017)**, which can result from the disruption of a class of germline-specific small RNAs known as Piwi-interacting RNAs **(Specchia et al. 2010; Gangaraju et al. 2011)**. Following stress-induced epigenetic changes, transposon activation or DNA deletions would generate somatic changes, which might ultimately become heritable via *de novo* germline mutations **(Fanti et al. 2017)**. Studying the environmental sensitivity of developmental processes across different and evolutionary distant genetic systems offers a way to test the generalizability of these cryptic and *de novo* variation-based models, possibly providing fresh insights into a molecular basis of developmental plasticity. It also provides new knowledge on the pervasiveness of phenotypic plasticity and advances current understanding of the role that environmental induction plays in adaptive evolution.

Ciliated protozoans are a biological system that enables easy manipulation of environmental conditions during development. In ciliates, nuclear development and germline-soma differentiation take place within a single cell **(Sonneborn 1977; Prescott 1994)**. Early studies in the ciliate *Paramecium* have shown that the exposure of genetically identical cells to different environmental conditions during nuclear

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

differentiation, the so-called 'sensitive period', leads to heritable phenotypic variations **(Jollos 1921).** Two of these environmentally sensitive traits have been extensively characterized. The first is the A system of complementary mating type determination, wherein the temperature to which clonal *Paramecium* cells are exposed during the sensitive period greatly influences mating type expression **(Sonneborn 1947)**. The second is the trichocyst discharge phenotype in homozygous clones of *P. tetraurelia*, where both temperature and food availability during development radically affect the phenotype expressed **(Sonneborn and M.V. 1979)**. The stable trans-generational epigenetic inheritance of phenotypes is well established in *Paramecium*, leaving open the possibility that environmentally sensitive traits may have evolutionary consequences in this ciliate. Furthermore, the mechanisms behind non-Mendelian trait inheritance in *Paramecium* are beginning to be understood at the molecular level **(Duharcourt et al. 1998; Garnier et al. 2004; Lepere et al. 2008; Duharcourt et al. 2009; Singh et al. 2014)**. Specifically, an intricate small RNA-mediated trans-nuclear crosstalk allows at least part of the genetic variability in the somatic nucleus to be inherited trans-generationally **(Coyne et al. 2012; Allen and Nowacki 2017)**. This knowledge is salient to investigations aimed at exploring the evolutionary impact that environmental changes may have on germline-soma differentiation in *Paramecium*.

Nuclear development in ciliates is coupled with a spectacular, reproducible process of selective DNA elimination from the developing somatic genome commonly known as Programmed DNA Elimination (PDE). PDE has been thoroughly characterized in *P. tetraurelia*—in addition to a broad range of eukaryotes **(Wang and Davis 2014)**, such as sea lamprey **(Smith et al. 2018b),** finches **(Biederman et al. 2018)** or humans **(Jung et al. 2006)**. The germline genome of this ciliated protozoan comprises some 45,000 mainly unique sequences known as Internal Eliminated Sequences (IESs) **(Arnaiz et al. 2012)**. These intervening sequences are flanked by two 5'-TA-3' dinucleotides whose disruption causes IES retention **(Mayer and Forney 1999)**, and reside both nearby and within genes, often interrupting open reading frames (ORFs). At each event of sexual reproduction, *P. tetraurelia* undergoes nuclear replacement *i.e.* the maternal somatic macronucleus is degraded and new macronuclei are produced through amplification (from 2n to ~800n) and extensive rearrangement of the germline genome housed in mitotic copies of the zygotic nucleus **(Betermier and Duharcourt 2014)**. At this stage, IESs must be reproducibly eliminated from the germline template; their accurate and efficient splicing from genes is essential for the correct functioning of the somatic genome and the production of viable sexual offspring **(Arnaiz et al. 2012)**. Lethal developmental defects result when Piggy MAC (PGM), the domesticated transposase required for the excision of virtually all IESs and forming complexes with five additional partners **(Bischerour et al. 2018)**,

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is silenced **(Baudry et al. 2009; Dubois et al. 2012)**. However, viable sexual offspring are yielded when Dcl2 and Dcl3, dicer-like proteins involved in the biogenesis of small RNAs (scnRNAs) and in the excision of a subset of IESs **(Lepere et al. 2009; Sandoval et al. 2014; Hoehener et al. 2018)**, are (independently) RNAi silenced. This latter observation demonstrates that inefficient IES excision can be tolerated to some degree. Additionally, two types of erroneous DNA elimination under spontaneous conditions (in addition to variable chromosome fragmentation) were described even before IESs were comprehensively catalogued **(Duret et al. 2008; Catania et al. 2013)**: Erroneous IES excision and 'cryptic IES recognition'. The former consists largely of inefficient excision, where IESs are excised from only a fraction of the ~800 macronuclear copies, and to a lesser extent, of alternative or nested boundaries usage. The second type prompts the elimination of IES-like bits of somatic DNA. Therefore, although PDE is a largely reproducible developmental program for the exclusion of unwanted DNA from the somatic lineage, inaccurate DNA elimination may take place during nuclear differentiation, leading to the production of alternatively rearranged versions of the same genome.

The somatic variability that (an inefficient) PDE introduces in otherwise genetically identical *Paramecium* cells **(Caron 1992; Duret et al. 2008)** can be the basis for at least part of the phenotypic differentiation in identical clones. For example, in *P. tetraurelia* and *P. octaurelia*, an IES-like somatic region (cryptic IES) containing the promoter and the transcription start site of the mtA gene is variably spliced during development, resulting in complementary mating type determination **(Singh et al. 2014)**. A similar mechanism is found in *P. septaurelia*, a closely related species, where a cryptic IES is removed from the coding region of mtB, a putative transcription factor required for the expression of mtA, leading to mating type switch **(Singh et al. 2014)**. Thus, PDE in *Paramecium* has been repeatedly coopted for the regulation of gene expression through alternative DNA splicing **(Orias et al. 2017)**. It is conceivable that DNA-splicing-controlled phenotypes in *Paramecium* have evolved via selection of heritable alternative DNA splicing variants, consistent with previously proposed models of epigenetic evolution **(Coyne et al. 2012; Allen and Nowacki 2017)**. Although cryptic IES excision is thus far the only characterized mechanism of PDE-dependent phenotypic diversification, in principle other sources of somatic variability such as inefficient IES excision could contribute to the emergence of genetic novelties **(Catania et al. 2013; Catania and Schmitz 2015)** and adaptive phenotypic plasticity **(Noto and Mochizuki 2017; Noto and Mochizuki 2018)**.

In this study, we tested the effect that the environmental temperature has on germline-soma differentiation in *P. tetraurelia*. Our findings demonstrate, for the first time, that programmed DNA elimination in ciliates is an environmentally sensitive

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

process. Since a large number of the IESs affected by temperature changes are epigenetically controlled and are passed down to sexual offspring, our findings also indicate that PDE is a powerful molecular ‘stonecutter’ capable of generating adaptive somatic variability.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Temperature affects the rate of incomplete IES excision and cryptic IES recognition

After allowing isogenic lines of *P. tetraurelia* to undergo autogamy (self-fertilization) at three different temperatures, 18˚C, 25˚C, and 32˚C, we inspected the independently rearranged somatic genomes of these lines (18˚C_{F1}, 25˚C_{F1}, and 32˚C_{F1}) and their progenitor (25˚C_{F0}) for incomplete IES excision and cryptic IES recognition. Hereinafter we will refer to 18˚C and 32°C as suboptimal temperatures compared to 25°C. We also arbitrarily define IES Retention Scores (IRSs) > 0.1 as non-trivial.

We detected ~400 IES loci with IRS > 0.1 (hereinafter, somatic IESs) in the progenitor line 25˚C_{F0} (**Figure 1A**). This count is comparable to that estimated for the descendant 25˚C_{F1} and remains similar between the two 25˚C samples, but it changes with the IRS threshold applied—note that the IRS is unaffected by between-sample variation in read coverage. In contrast, the count of somatic IESs is considerably higher in the macronuclear genomes that developed at suboptimal temperatures. We detected up to ~800 somatic IESs in macronuclear genomes rearranged at 18°C and 32°C, roughly a two-fold increase compared to the optimal temperature. The vast majority of these somatic IESs have scores in the range of 0.1-0.3 (**Figure 1A**), with a potential impact on gene expression.

Suboptimal temperatures affect also the rate of cryptic IES recognition (TA-bound somatic deletions) (**Figure 1B**). We detect up to a ~2-fold increase when comparing the number of partially excised cryptic IESs unique to 18˚C and 32˚C with those unique to the 25˚C samples (**Figure S1**). Many TA-bound somatic deletions lead to partial or even complete gene ablation and occasionally span multiple genes at once (a catalogue is presented in **Table S1)**. Among the 18 TA-bound somatic deletions consistently retrieved across all temperatures (maternally inherited), we find the 195 bp-DNA segment containing the promoter and transcription start site of mtA, a DNA-splicing regulated gene **(Singh et al. 2014; Orias et al. 2017)**. We also report a set of IES-like somatic regions that are variably spliced at different temperatures, which might represent temperature-sensitive cryptic IESs. The surge in cryptic IES deletions with temperature is less pronounced compared to that in the annotated IESs and mostly limited to deletion scores (DS) that tend to be smaller than 0.1 (**Figure 1B**). As a consequence, we decided to focus the presentation of our results on true IESs.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

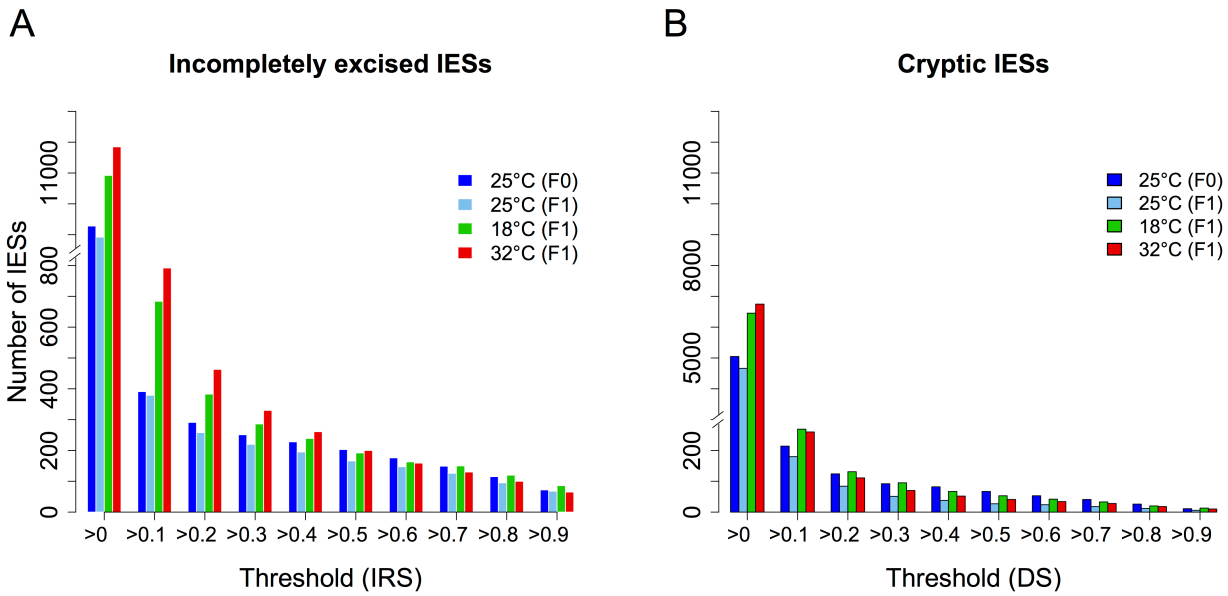


Figure 1. The rate of incomplete IES excision and somatic DNA deletion is temperature dependent. (A) The number of detected incompletely excised IESs is higher at 18°C and 32°C compared to 25°C. For non-trivial IES retentions (IRS > 0.1), the mean count bottoms at ~400 IESs at 25°C, whereas it rises up to ~750-800 IESs at suboptimal growth temperatures (18°C and 32°C). (B) TA-bound somatic deletions (Cryptic IESs) are also more frequent when PDE occurs at suboptimal growth temperatures. IESs, Internal Eliminated Sequences; PDE, Programmed DNA Elimination; IRS, IES Retention Score; DS, Deletion Score.

Suboptimal environmental temperatures decrease PDE efficiency

We investigated how extensively temperature changes affect the magnitude of IES retention in the polyploid somatic genome. The full set of PGM-IESs with their retention scores (IRS), and the genome-wide testing of the F0-to-F1 IRS transitions for all samples are reported in **Table S2**.

We detected a marked reduction in PDE efficiency at 18˚C and 32˚C relative to 25˚C (**Figure 2**). The number of IESs with a greatly intensified retention in the F1 somatic nuclei ($IRS_{F1} > IRS_{F0}$, Binomial test, $P_{adj} < 0.05$) rises from 43 at 25˚C, to 183 and 271 at 18˚C and 32˚C, respectively. Further, most of the significantly retained IESs detected at 18˚C and 32˚C are unique to sub-optimal temperatures—with a treatment-control ratio of ~12-fold (151:13) and ~17-fold (225:13) for 18°C and 32°C, respectively (**Figure S2A**). Yet the number of IESs excised with significantly *increased* efficiency in the F1 generation is comparable across temperatures (**Figure 2**), with around 50% overlap between each of the experimental lines and the control (**Figure S2B**). The count of significant transitions for the three temperatures tested

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are summarized in **Figure 2D**. The count ratio of upward (reduced excision efficiency) to downward (increased excision efficiency) IRS transitions after sexual reproduction approaches 1 (43:57) for the control temperature of 25˚C, whereas it rises up to ~3.5 (183:54) and ~5.5 (271:50) at 18°C and 32°C, respectively. These observations demonstrate that rather than producing stochastic effects, in our study suboptimal temperatures substantially reduce PDE efficiency, with the rate of erroneous excision rising well above biological noise.

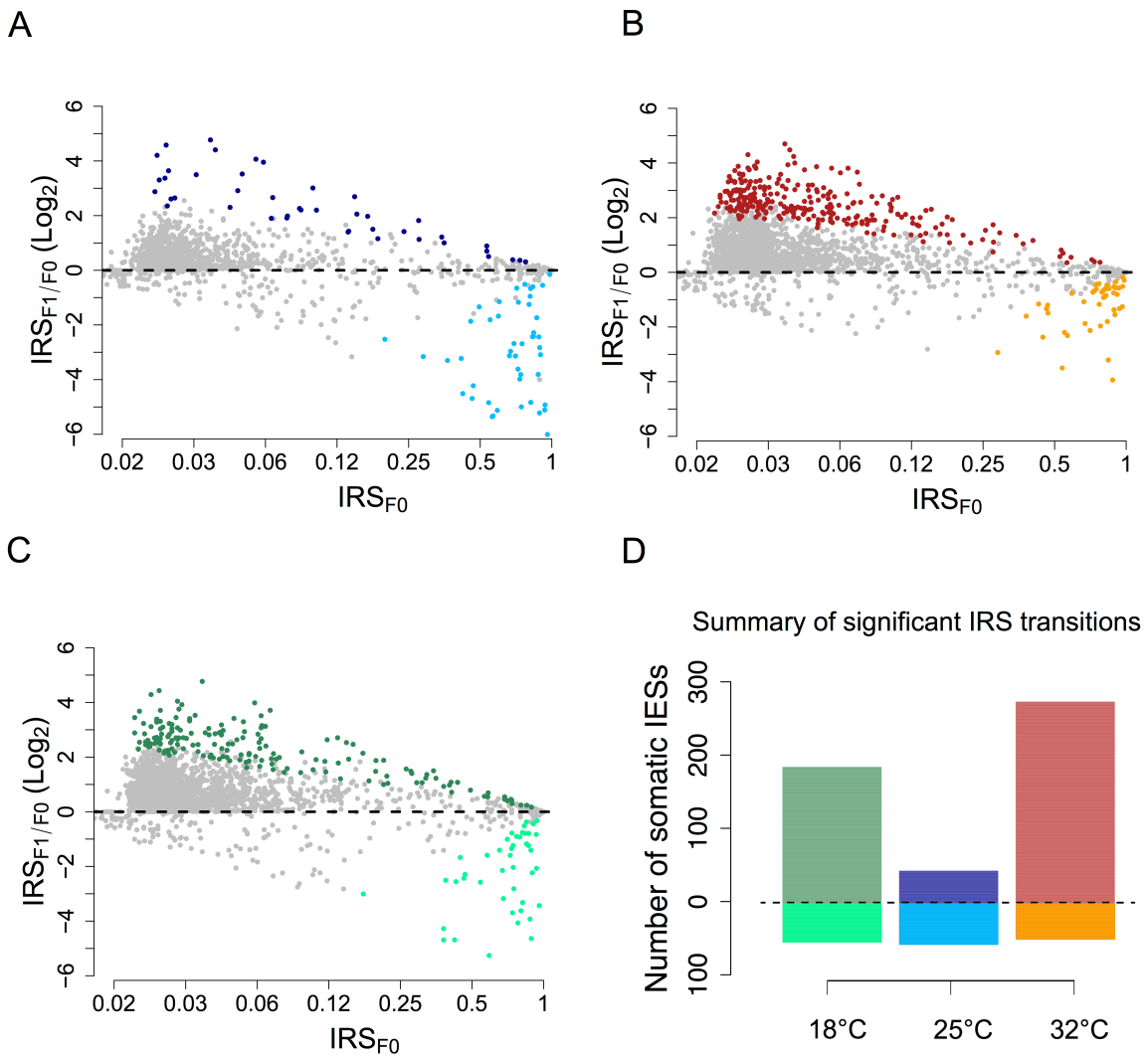


Figure 2. PDE inefficiency shows a characteristic ∪-shaped relationship with temperature. (A-C) Bland-Altman plots displaying the Log2 fold change of IRSs from F0 to F1 for 25˚C, 32˚C and 18˚C, respectively. Statistically significant IRS transitions (Binomial test, $P_{adj} < 0.05$) are shown as colored filled-circles: 25˚C, dark and light blue circles; 32˚C, red and orange circles; 18˚C, dark and light green circles (dark color: $IRS_{F1} > IRS_{F0}$; light color: $IRS_{F1} < IRS_{F0}$). Position on x-axis reflects the log-transformed initial (F0) state of the IRSs (x-labels are IRSs before log transformation). (D) Counts of

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

statistically significant IRS transitions (P_{adj} < 0.05) after nuclear differentiation at 18˚C (green), 25˚C (blue) and 32˚C (red). The number of somatic IESs experiencing an upward or downward IRS transition is shown above and below the horizontal dashed line, respectively.

Incomplete IES excision is trans-generationally inherited

Until now, we have mainly focused on the performance of PDE at different environmental temperatures. We now turn to the potential biological significance of the PDE-mediated molecular variation. To begin, we asked whether the somatic IESs detected in the F1 genomes were all generated anew or if they were partly obtained through trans-somatic inheritance.

To gain insight into this question, we assessed how many IESs are retained simultaneously in the four independently rearranged F0 and F1 somatic nuclei. If incomplete IES excisions are truly stochastic events as commonly regarded, then the observed number of 4-way-shared IESs should not exceed what would be expected by chance. Leveraging 100 simulated datasets based on random draws from the PGM-set (*i.e.*, the set of IESs retained after Piggy MAC silencing), we estimated that the expected maximum number of IESs shared among four genomes is ~107. We observed 934 4-way-shared IESs, a striking ~ninefold increase (**Figure 3A**).

In principle, somatic IESs shared between independent macronuclei could reflect weak *cis*-acting IES recognition/excision signals. Weak splicing signals might explain the significantly elevated median IRS (up to complete retention, IRS ~ 1) of the 4-way-shared IESs relative to the full-set of incompletely excised IESs illustrated in **Figure 3B**. Consistent with the weak-signal hypothesis is the study of the C_{in}-score—a predictor of splicing signal quality **(Ferro et al. 2015)**—which reveals statistically smaller C_{in}-score estimates for 4-way-shared IESs relative to the PGM-set ($IESs_{4\text{-}way}$ vs. $IESs_{PGM}$, 0.54 vs 0.62, Mann-Whitney U, P < 0.001). This difference holds true when we control for size class and genomic location. Nonetheless, weak splicing signals alone cannot explain why the same IESs exhibit significantly elevated median IRS at 32°C and 18°C relative to 25°C (**Figure 3B**).

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

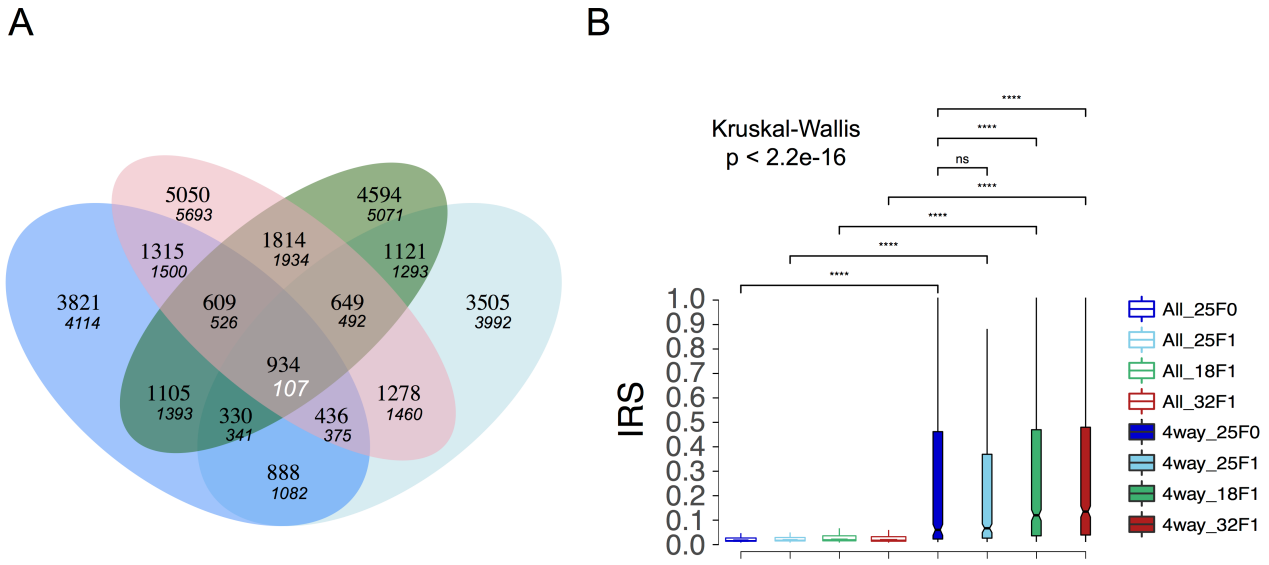


Figure 3. Incompletely excised IESs are inherited trans-generationally and show elevated retention scores (A) Venn diagram depicting sets of incompletely excised IESs shared between three F1 somatic genomes and their parental F0 genome. A large excess of incompletely excised IESs is shared across all four genomes, indicating that most of the IESs in the 4-way set were inherited trans-somatically from F0 to F1. The maximum number of shared IESs, as determined over 100 random draws, is shown under each observed overlap value. Blue ellipse, 25˚C (F0), green ellipse, 18˚C (F1), light blue ellipse, 25˚C (F1), red ellipse, 32˚C (F1). (B) Box plot showing the IRS distributions (IRS > 0) for the full-set and the 4-way set (4w) of incompletely excised IESs at all investigated temperatures. Pairwise comparisons between groups were performed using a Wilcoxon Rank Sum Test with correction for multiple testing (BH, Benjamini–Hochberg). Statistical significance is indicated for each comparison (****; $P < 0.001$, ns, non-significant). Outliers are omitted for clarity.

Another (not mutually exclusive) explanation for the excess of somatic IESs common to subsequent sexual generations is that these somatic IESs might reflect episodes of possibly ongoing trans-generational epigenetic inheritance. Under these circumstances, we expect that many of the discussed somatic IESs (including the 4-way-shared IESs) be epigenetically regulated. We tested this hypothesis taking advantage of published knock down (KD) studies of PDE-associated epigenetic components in *P. tetraurelia* **(Sandoval et al. 2014)**. Three Dicer-like endonucleases and two classes of developmentally specific small RNAs were shown to be necessary for the accurate excision of a few thousands of *P. tetraurelia* IESs: Dcl2 and Dcl3 are required for the biogenesis of scnRNAs in the germline nucleus **(Lepere et al. 2009)**, whereas Dcl5 is responsible for the production of iesRNAs in the developing somatic nucleus **(Sandoval et al. 2014)**. First we examined the set(s) of somatic IESs that are significantly retained, without being necessarily shared between independent

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

macronuclei. We found that the excision of roughly half of these IESs is dependent on Dcl2/3 and or Dcl5, *i.e.*, their retention score increases significantly upon Dcl2/3 and Dcl5 KD (**Figure 4**).

We then examined how extensively the 4-way-shared IESs are under the control of Dcl2/3-5 (data taken from ParameciumDB). We found that 166 (out of the 934) IESs are indeed Dcl2/3-controlled IESs and 248 are Dcl5-controlled. This establishes that at least 35% of the 4-way-shared somatic IESs are under epigenetic control, a ~3.6-fold enrichment compared to random expectation (~10%), and thus likely to be epigenetically inherited.

In fact, the number of 4-way-shared incompletely excised IESs that are epigenetically controlled/inherited might be even larger. Under the scnRNA model of trans-nuclear comparison **(Duharcourt et al. 2009)**, IESs that are severely retained in the maternal macronucleus will be less affected by the scnRNA depletion ensuing Dcl2/3 KD compared to IESs that are absent from the maternal macronucleus. Conventional approaches may fail to classify these highly retained IESs as Dcl2/3-dependent because the shifts between pre- and post-KD levels of IES retention might be negligible. Leveraging the IRSs obtained by previous Dcl2/3 KD experiments **(Lhuillier-Akakpo et al. 2014; Sandoval et al. 2014)** we identified 236 IESs that despite having particularly elevated IRSs in these KD experiments (IRS > 0.3) were not recorded as influenced by the scnRNA machinery. Around 68% of these IESs (n=160) are found in our 4-way-shared set. Thus, our approach based on perturbation of environmental conditions permits the identification of trans-somatically inherited IESs without requiring any previous knowledge of the epigenetic factors involved in their excision, adding hundreds of candidate epigenetically-controlled IESs to the existing set.

Collectively, our observations suggest that a considerable number of somatic IESs in *P. tetraurelia* are trans-somatically inherited, and that the retention levels of these inherited IESs are enhanced in the sexual progeny upon exposure to suboptimal temperatures.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

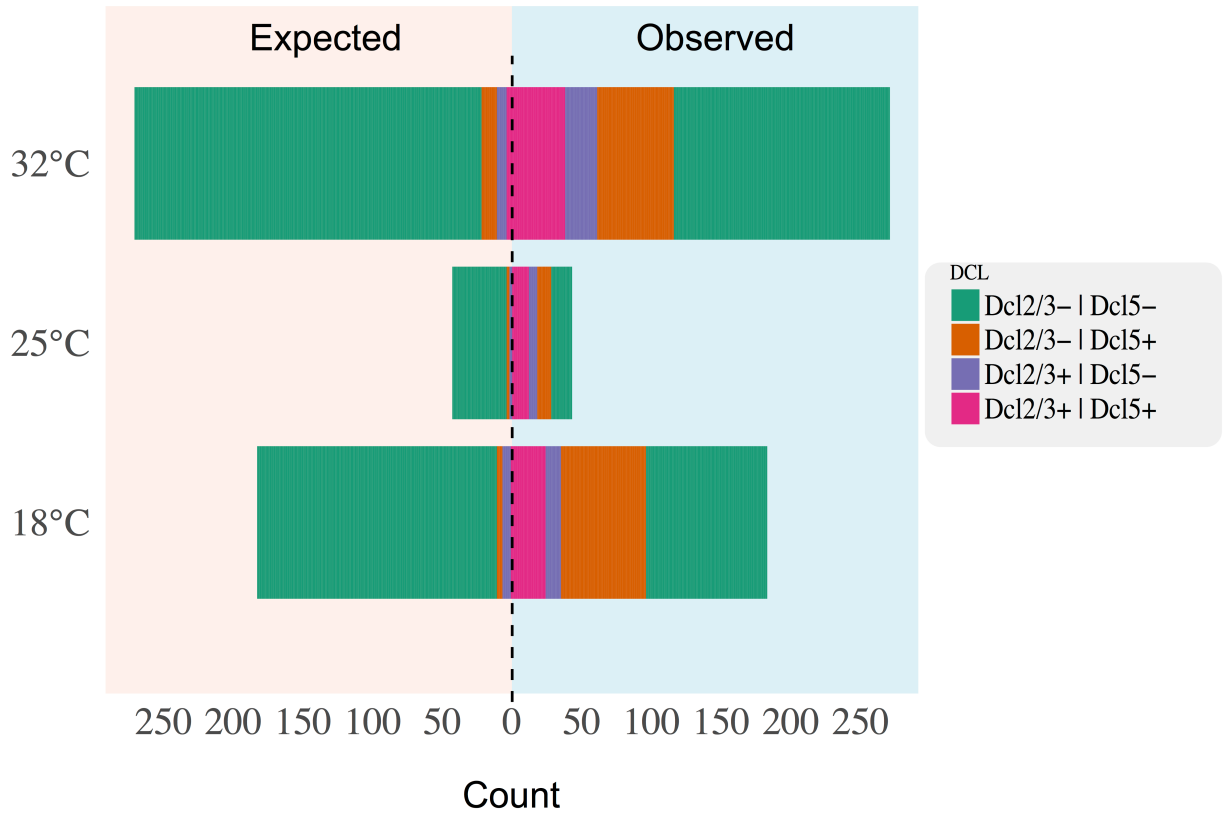


Figure 4. A large excess of incompletely spliced IESs is epigenetically regulated. Back to back stacked bar chart showing the number of significantly retained IESs after nuclear differentiation at all investigated temperatures. For each temperature, IES counts are broken down into Dcl2/3-controlled IESs (Dcl2/3^{+} | Dcl5^{-}, Purple), Dcl5-controlled IESs (Dcl2/3^{-} | Dcl5^{+}, Orange), Dcl2/3-Dcl5-co-controlled IESs (Dcl2/3^{+} | Dcl5^{+}, Fuchsia) and Dcl-independent IESs (Dcl2/3^{-} | Dcl5^{-}, green). Expected proportion of Dcl-dependent IESs for random samples of the same size (left side) are shown back to back with the observed data (right side).

Purifying selection shapes IES retention profiles

To deepen our understanding of the biological significance of somatic IESs, we performed a systematic analysis of their genomic distribution. The expectation is that non-trivially retained IESs are more prevalent in i) intergenic regions and ii) weakly expressed genes or genes that are not critical for development or cell viability.

Indeed somatic IESs (IRS > 0.1) preferentially occupy intergenic regions (**Figure 5A**). Furthermore, they are more likely to occur in genes that in the *P. tetraurelia* strain 51 are weakly expressed in the vegetative stage (**Figure 5B**). Interestingly, the ratio of intergenic to intragenic somatic IESs is IRS-dependent. The proportion of intergenic somatic IESs increases abruptly as the IRS crosses 0.1 (**Figure 5A**) and plateaus at

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

~70% beyond an IRS threshold of ~0.25. This is true at all the investigated temperatures.

When we compared the subset of significantly retained IESs (see *Suboptimal environmental temperatures decrease PDE efficiency*) with the PGM-set, we once again detected an excess of intergenic loci (**Figure 5C**). Nevertheless, a cross-sample comparison reveals a substantial enrichment of somatic IESs within exons at 18°C and 32°C relative to 25°C (**Figure 5C**). In particular, 83 and 158 genes show significant levels of IES retention at 18˚C and 32˚C, respectively, compared to only 17 genes at 25˚C (**Figure S3**). Intriguingly, the deviation from the reference distribution is for the most part determined by the 4-way-shared set of IESs. After the exclusion of these largely epigenetically-controlled IESs, the genomic distribution of the remaining set of retained IESs conforms to the expected reference distribution (data not shown).

In sum, although IESs in intergenic regions are generally more likely to be incompletely excised, suboptimal environmental temperatures appear to mostly perturb the excision of exonic, epigenetically regulated, and presumably trans-somatically inherited IESs.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

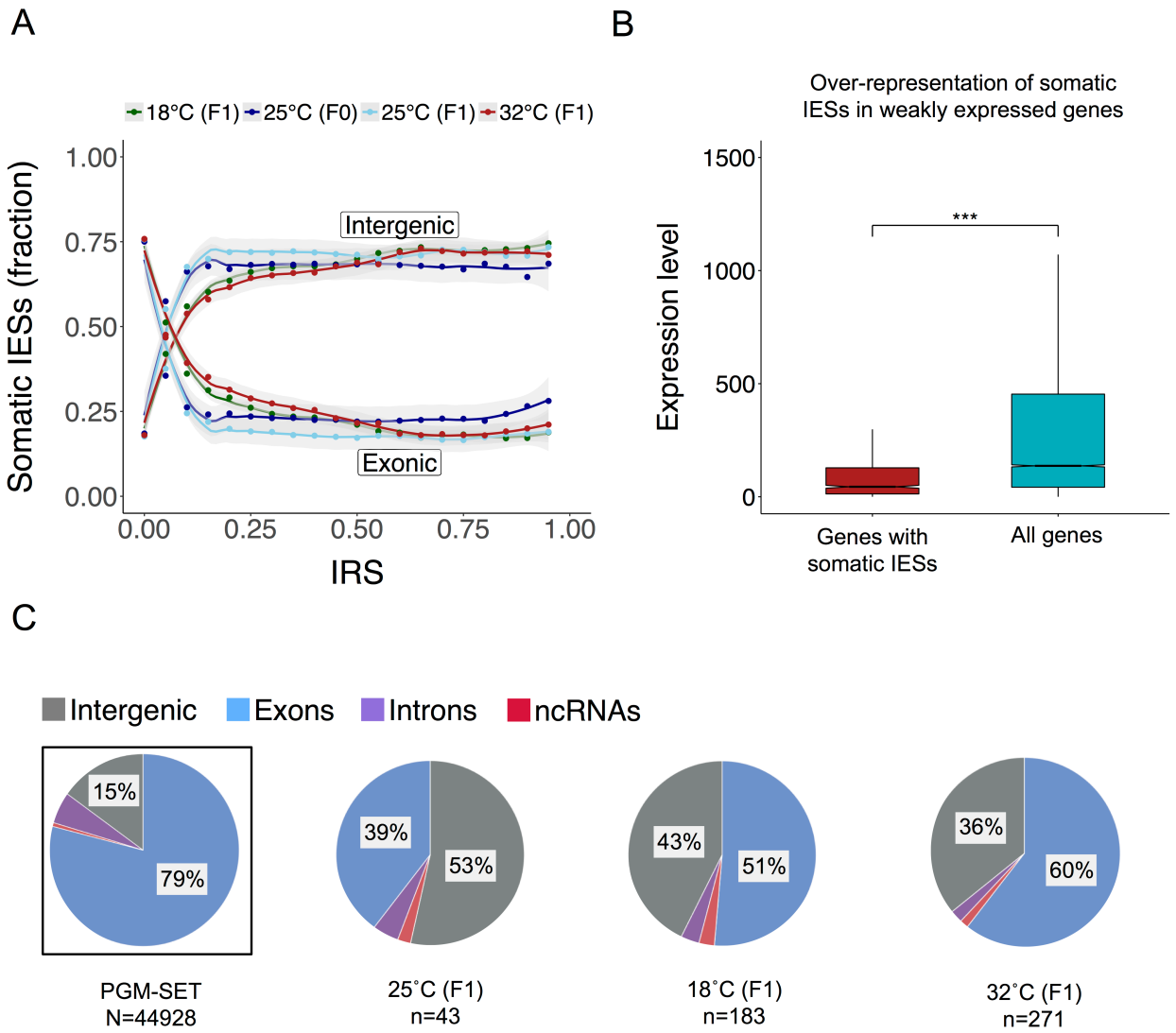


Figure 5. Non-trivially retained IESs are under purifying selection. (A) IRS-dependent changes of nuclear prevalence for incompletely excised intergenic and exonic IESs. (B) Expression levels distributions of genes affected by IES retention (IRS > 0.1, red box) and the full set of *P. tetraurelia* macronuclear coding genes (cyan box). Pairwise comparison was performed with a Mann–Whitney *U* test. Statistical significance is indicated (***; $P < 0.01$). (C) Genomic distribution of significantly retained IESs at all investigated temperatures. Distribution of PGM-controlled IESs is shown for reference (leftmost pie chart, boxed). Sample size is indicated in brackets. Percentages below 5% are not indicated.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

IES retention in exons disrupts ORFs largely, but not exclusively

Somatic IESs within genes could give rise to functional alternative isoforms. But how likely is this event to occur? Within coding sequences (CDS), IES retention can induce Premature Termination Codons (PTCs), either through within-IES PTC (IES-PTC) or more commonly via PTC induction downstream to the inserted element (FrameShift-PTC or FS-PTC), likely contributing to transcript degradation via Nonsense-mediated mRNA Decay (NMD) **(Brogna and Wen 2009).** In addition, IES insertion near the 3' end of the CDS can lead to the ablation of the true stop codon (Tail-FS IESs), presumably resulting in mRNA degradation via non-stop mediated RNA decay **(Frischmeyer et al. 2002; Vasudevan et al. 2002; Klauer and van Hoof 2012)**. Nevertheless, productive alternative DNA splicing variants could, at least in theory, be achieved via retention of non-PTC containing 3n-IESs (length multiple of 3).

To assess the impact of incomplete IES excision on genes, we calculated the occurrence of PTC-inducing and non-PTC inducing IESs in the PGM-set (control set) and compared it with their counterpart in the experimental 32˚C set (which contains the largest number of significantly retained IESs). We find that ~20% of IES retentions may in theory produce protein diversification, although only ~5% of the IESs retained in the experimental set represent cases of theoretically productive alternative DNA splicing (3n-IESs; **Figure 6**). Of note, Tail-FS IESs outnumber 3n-IESs in the 31 genes of the experimental set with *bona fide* CDS extension (15.8% vs. 5.4%), whereas the reverse pattern is found for the control set (3.3% vs. 20.6%). We infer that the vast majority of incomplete IES excisions would likely impact protein availability—a condition that by silencing some genes rather than others might advantageously facilitate adaptation to a new environment—although cases of potentially productive, IES-driven protein diversification may occur at each event of sexual reproduction.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

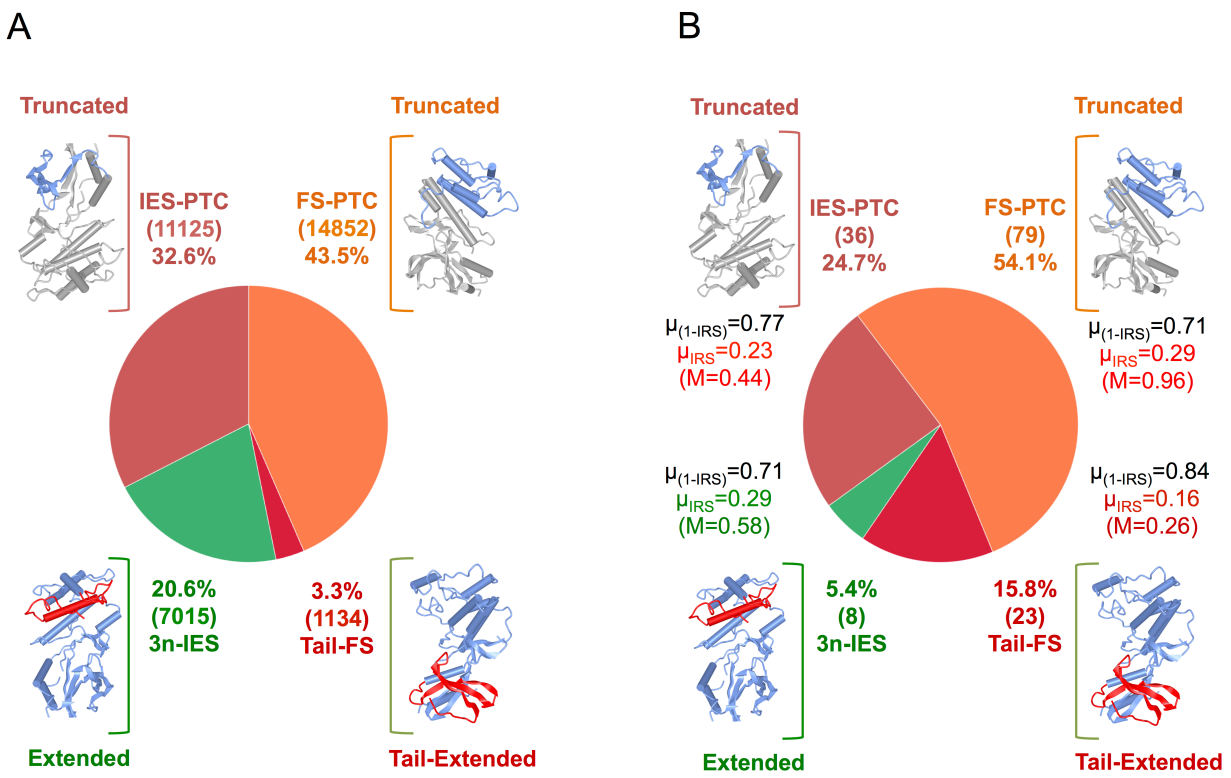


Figure 6. IES insertion within CDS potentially diversifies protein sequences. (A) Theoretical diversification potential of PGM-IESs. 3n-IES (green): fraction of productive, non-PTC inducing IESs. Tail-FS (crimson red): fraction of IESs that do not introduce a PTC but lead to the ablation of the true stop codon. FS-PTC (orange): fraction of IESs that disrupt the open reading frame by introducing a downstream PTC. IES-PTC: Fraction of IESs that introduces a PTC within the insertion. (B) Classification of the 146 within-CDS IES insertions observed at 32˚C based on the predicted transcriptional outcome. Mean excision score (1-μIRS), Mean IRS (μIRS), Maximum IRS (M) are indicated next to each category. The number of IESs is indicated within brackets. The predicted effect on the protein products is depicted schematically next to each class. IES, Internal Eliminated Sequence; CDS, Coding Sequence; PTC, Premature Termination Codon; PGM, Piggy Mac transposase; FS, Frame-Shift.

Non-random distribution of IESs with respect to protein families and molecular function

Finally, we explored possible biases in the molecular function of the genes with significant IES retention in coding sequences. The functional categorization of these genes is given in **Figure S3**. The full set of genes, their expression values (as in strain 51; taken from **(Arnaiz et al. 2017)**), and annotations, along with parameters related to the retained IESs are presented in **Table S2**.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A GO term enrichment analysis reveals a single molecular function term, *protein binding*, enriched at both 18˚C and 32˚C (Fisher exact test, $P < 0.0001$), but not at 25˚C ($P = 0.05$). Interestingly, two sets of proteins, the Tetratrico Peptide Repeat Region (TPR, IPR019734) and Growth Factor Receptor cysteine-rich domain (GFR, IPR009030) containing proteins contribute largely to this functional enrichment. More specifically, TPR and GFR-genes together account for ~21% and ~24% of the genes affected at 32˚C and 18˚C, respectively. This indicates that in *P. tetraurelia* there may be genes that are more susceptible to incomplete IES excision than others.

We hypothesized that there are genes in *P. tetraurelia* that are particularly IES-dense and thus more likely to be among the genes hit by IES retention. In testing this hypothesis, we found that genes significantly affected by IES retention at 18˚C and 32˚C do have significantly greater than average number of IESs (**Figure 7A**) and IES density (**Figure 7B**). Furthermore, TPR motif-containing proteins (IPR019734) and GFR cysteine-rich domain-containing proteins (IPR009030) exhibit significantly elevated numbers of IESs *per* gene, being among the most IES-rich genes in the *P. tetraurelia* genome (**Figure 7C**). While both the GFR and TPR protein families are extremely IES-rich, only the latter is also ultra-IES dense (**Figure 7D**): with IES densities up to 10 IES/kb, the TPR-motif family of proteins alone accounts for almost 3% (1,200 IESs) of the 44,928 PGM-set of IESs (an example of TPR-motif gene is in **Figure 7E**).

The pronounced representation of GFR- and TPR-containing proteins in our dataset might be merely expected by chance. To address this question, we partitioned *P. tetraurelia* genes into three groups on the basis of their InterPro domain annotations, TPR, GFR and Protein kinase-like domain (PKD) genes, with the latter group serving as a control in the enrichment analysis. We find that the number of GFR-genes (as well as PDK-genes) does not differ significantly from the expected values (see Materials and Methods), neither at 32˚C nor at 18˚C (binomial test, $P > 0.05$). Conversely, TPR-genes are highly overrepresented at both sub-optimal temperatures (binomial test, $P < 0.0001$). Thus, TPR-genes in *P. tetraurelia* appear to be highly susceptible to IES retention at sub-optimal temperatures.

Guided by this non-random distribution of IESs in relation to protein families, we next asked whether *P. tetraurelia* IESs are generally non-randomly distributed with respect to molecular functions or biological processes. To address this question, we performed enrichment analysis on the subset of *P. tetraurelia* IES-containing genes using the full set of *P. tetraurelia* macronuclear genes as reference. Remarkably, 10 molecular function and 8 biological process terms are enriched in the set of IES-containing genes (**Table 1**). This suggests that *P. tetraurelia* genes involved in

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

specific cellular functions such as ion transport, signal transduction and microtubule-based movement, among others, are relatively more likely to contain IESs. Conversely, we find a deficit of IESs in e.g. translation-associated factors.

These findings demonstrate that IESs are not randomly distributed in relation to protein families, molecular functions, and biological processes. The functional enrichment of IES-containing genes could be explained in terms of i) a spatially patterned genomic invasion by transposable IES progenitors due to heterogeneous levels of purifying selection that antagonizes gene interruption and/or ii) differential expansion of gene families in the *P. tetraurelia*'s genome following the invasion. An alternative, tempting explanation is that IESs' patterned genomic topography has been shaped by natural selection over evolutionary time to help regulate organismal responses to environmental changes.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

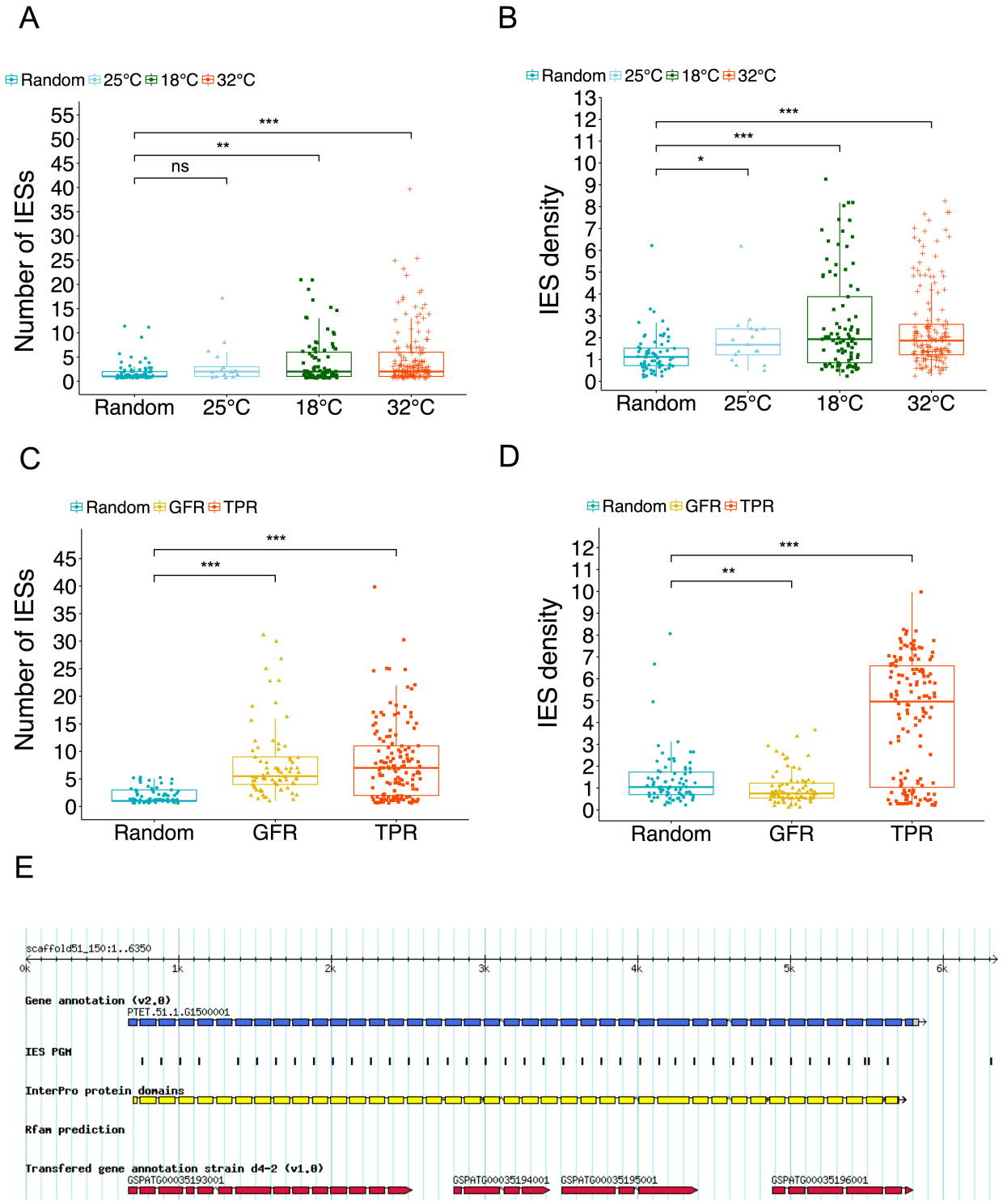


Figure 7. TPR-motif and GFR-Cys-rich domain containing proteins are IES-rich gene families susceptible to inefficient IES excision in response to temperature changes during autogamy. (A) Genes affected by IES retention at 18˚C and 32˚C have significantly greater than average number of IESs. (B) Across all the investigated temperatures, genes hit by significant IES retention are IES-dense (Kruskal-Wallis, $P < 0.0001$). (C) TPR and GFR proteins exhibit extraordinary per gene IES

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

counts (Kruskal-Wallis, $P < 0.0001$). (D) TPR but not GFR proteins are characterized by elevated IES density (Kruskal-Wallis, $P < 0.0001$). (E) A TPR-motif gene (PTET.51.1.G1500001) showing a characteristic pattern of IES distribution. This gene is hit by multiple IES retention at 32˚C. IESs are positioned almost invariably at the 5' end of the TPR coding exons. According to the v2.0 annotation of macronuclear gene models, PTET.51.1.G1500001 is the gene with the greatest number of IESs to be found in the *P. tetraurelia* genome. TPR, Tetratrico Peptide Repeat region motif (IPR019734). GFR-Cys-rich, Growth Factor Receptor cysteine-rich domain (IPR009030). IES, Internal Eliminated Sequence. IES density, number of IESs per kb.

Table 1. **IES-containing genes are enriched in specific molecular functions and biological processes.** GO ID, GO term IDs. Annotated, number of genes mapped to the corresponding GO term in the genome. Obs, Observed number of genes; Exp, Expected number of genes. Fold, Fold Enrichment of the functional category. FDR (False Discovery Rate), Fisher's exact test with Benjamini-Hochberg correction for multiple testing, $P_{adj} < 0.05$. The results shown in the table were obtained using the *Panther gene list analysis* tool. Similar results were obtained performing the functional enrichment analysis with the *topGO* package (not shown). The outcome of the GO-Term analysis was highly consistent between gene annotation versions (macronuclear gene models v1 and v2).

GO ID	Molecular Function	Annotated	Obs	Exp	Fold	FDR
GO:0016849	phosphorus-oxygen lyase activity	63	56	29.96	1.87	1.25E-02
GO:0008081	phosphoric diester hydrolase activity	73	59	34.72	1.70	3.97E-02
GO:0005249	voltage-gated potassium channel activity	78	63	37.10	1.70	3.02E-02
GO:0000155	phosphorelay sensor kinase activity	177	142	84.18	1.69	1.12E-04
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	144	104	68.49	1.52	2.03E-02
GO:0003777	microtubule motor activity	245	174	116.52	1.49	1.34E-03
GO:0005524	ATP binding	3777	2300	1796.37	1.28	3.14E-20
GO:0008270	zinc ion binding	511	305	243.04	1.25	2.60E-02
GO:0070011	peptidase activity	639	378	303.91	1.24	1.25E-02
GO:0004674	protein serine/threonine kinase activity	962	560	457.54	1.22	2.65E-03
GO:0003924	GTPase activity	386	138	183.59	0.75	4.92E-02
GO:0003735	structural constituent of ribosome	404	100	192.15	0.52	1.97E-08
GO ID	**Biological Process**	**Annotated**	**Obs**	**Exp**	**Fold**	**FDR**
GO:0009190	cyclic nucleotide biosynthetic process	63	56	29.96	1.87	2.52E-02
GO:0006816	calcium ion transport	66	57	31.39	1.82	3.82E-02
GO:0006813	potassium ion transport	100	77	47.56	1.62	4.66E-02
GO:0000160	phosphorelay signal transduction system	224	165	106.54	1.55	1.10E-03
GO:0023014	signal transduction by protein phosphorylation	228	166	108.44	1.53	1.55E-03
GO:0007018	microtubule-based movement	269	185	127.94	1.45	5.67E-03
GO:0098662	inorganic cation transmembrane transport	268	181	127.46	1.42	1.06E-02
GO:0006508	proteolysis	691	404	328.65	1.23	3.01E-02
GO:0006412	translation	560	176	266.34	0.66	4.12E-05

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

In this work we asked: How extensively might environmental changes affect germline-soma differentiation? Answering this question contributes to our view of how the crosstalk between genes and environment affects organismal constitution and evolution. One possibility is that environmental changes may have generally little impact on this developmental process. Alternatively, environmentally induced perturbations might introduce considerable molecular variation into the developing somatic genome. Our results provide support for the second perspective. They also suggest that environmentally induced somatic variation in *Paramecium* might be evolutionarily relevant.

We show that the rate of spontaneous IES retention and cryptic IES recognition increase at sub-optimal temperatures (**Figure 1**), with the mean error rate of IES excision with IRS > 0.1 rising from ~200 at 25°C to ~600 erroneously excised IESs per sexual generation. In addition, the inefficacy of PDE during *P. tetraurelia*'s autogamy has a characteristic ∪-shaped relationship with temperature—the inefficiency of IES excision peaks at both low (18°C) and high temperatures (32°C) whereas PDE experiences an optimal performance at 25˚C (**Figure 2**). IES excision is therefore greatly sensitive to changes in the environmental temperature, a finding that may be surprising given that *Paramecium* is continuously exposed to quotidian and seasonal temperature fluctuations in naturally occurring conditions, thriving and undergoing sexual reproduction along latitudinal temperature gradients **(Krenek et al. 2011; Krenek et al. 2012)**. In as little as one sexual generation, the thermo-plasticity of IES excision translates into the introduction of hundreds of new alternative DNA splicing variants and in an elevated nuclear prevalence of standing somatic variation.

What is the cause of this increased somatic genetic variation? Intracellular processes coupled with DNA repair such as meiotic recombination have been recently shown to be thermo-plastic, with a similar ∪-shaped temperature-performance response **(Lloyd et al. 2018)**. Much like meiotic recombination, both biophysical and physiological alterations in response to temperature, such as protein-nucleic acid interactions and the oxidative state of the cell might contribute to the observed thermo-plasticity. That noted, we uncovered a sizeable excess of incompletely excised IESs that are epigenetically controlled (**Figure 3, Figure 4**). We therefore consider it most likely that the somatic variability introduced at sub-optimal temperatures depends significantly on the environmentally induced modulation or re-wiring of the epigenetic machinery regulating IES excision.

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Can the observed increment in somatic genetic variation have some biological relevance? The negative correlation between IRS and fraction of exonic IESs (**Figure 5**) indicates that natural purifying selection opposes the retention of exonic IESs. This relationship is of utmost importance: it makes it very likely that somatic variability has phenotypic consequences. This interpretation is consistent with published results concerning the quality of *cis*-acting IES recognition/excision signals—IESs with higher quality signals lie preferentially within, rather than outside of, genes **(Ferro et al. 2015)**. It is further strengthened by the observation that non-trivial IES retention events are mainly located within weakly expressed genes (**Figure 5**). Taken together, our findings demonstrate that part of the somatic variability in *P. tetraurelia* does not represent mere biological noise, but rather, biologically relevant selectable variation.

Are alternative somatic DNA splicing variants heritable? Our experimental setting allowed us to confidently capture trans-somatic inheritance in action. By leveraging the parallel sequencing of independently rearranged somatic genomes, we found that hundreds of somatic IESs, Dcl2/3- and Dcl5-controlled IESs, are very likely passed down to the sexual offspring after autogamy (**Figure 4**). Additionally, the nuclear prevalence of these somatic IESs increases in the sexual progeny upon exposure to suboptimal temperatures (**Figure 3B**). Thus, alternative DNA splicing variants may be heritable. Future studies evaluating the stability of trans-generational epigenetic inheritance of IESs across subsequent generations are required to determine e.g. whether mildly deleterious or potentially beneficial IES insertions can actually spread at the population level.

Is alternative somatic DNA splicing a source of functional innovation? To gain insight into this question, we first evaluated the impact of inefficient IES excision on gene sequences. Our observations suggest that increased IES retention following environmental perturbation result, in most cases, in the reduction of transcript availability, as inferred by the introduction of PTCs in the ORFs (**Figure 6**). In a small fraction of observed cases, IES insertion might additionally produce diversified protein sequences (**Figure 6**). Next, we performed an in-depth analysis of the genes hit by IES retention in response to PDE's thermo-plasticity. We found that at least one IES-rich gene family, TPR proteins, is particularly prone/susceptible to inefficient IES excision (**Figure 7**). Although the function of these proteins is currently unknown, their domain signatures suggest that they are involved in protein-protein interactions. Considering that TPR protein-coding genes are IES rich and yet successfully freed from IESs at 25˚C, it might simply be that the excision machinery performs particularly poorly in IES rich regions at sub-optimal temperatures. Alternatively, IES excision in TPR protein-coding genes may be actively modulated in sub-optimal environments as a mechanism of gene expression control and/or to facilitate protein diversification. This

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alternative hypothesis is consistent with the high enrichment of epigenetically controlled IESs, Dcl5-IESs in particular, in TPR protein-coding genes (not shown). Finally, we explored the possible impact of IES retention on cellular functions and found that IES-containing genes are significantly more likely to be involved in processes such as signal transduction, cellular protein modification, and transport of ions across membranes (**Table 1**). These results open the possibility that, much like the eukaryotic cellular process of alternative RNA splicing **(Lewis et al. 2003; Wong et al. 2013; Braunschweig et al. 2014; Marquez et al. 2015; Singh et al. 2017; Smith et al. 2018a)**, regulated IES retention in *Paramecium* might fine tune gene expression and/or generate alternative DNA splicing isoforms which may ultimately facilitate adaptation to environmental changes.

Our study offers fresh insights with regard to two current models—cryptic vs *de novo* variation-based—for explaining how selected traits become genetically encoded **(Kasinathan et al. 2017)**. Because *Paramecium* IESs are thought to be transposon-derived sequences **(Klobutcher and Herrick 1995)**, somatic IESs may be viewed both as *de novo* induced (somatic) insertions and unmasked cryptic (germline) variation at the same time. This ambiguity blurs the conventional distinction between cryptic and *de novo* induced genetic variation, rendering a discussion about the plausibility or generalizability of both models difficult. Instead, this ambiguity, together with intriguing parallels between our observations in *Paramecium* and previous findings in *Drosophila* concerning the non-random occurrence of transposon insertions in response to external stresses **(Jollos 1934; Fanti et al. 2017)**, enable cryptic and *de novo* induced variation-based mechanisms to be integrated. This exercise provides a comprehensive, potentially powerful framework for interpreting and predicting molecular dynamics associated with developmental plasticity across eukaryotes.

The framework proposed here is rooted in the idea that a number of cryptic genetic variants may be but classical genetic variants that while functional in an unstressed state, yield gene products with altered properties (e.g. down-regulated expression) in stressful environments. Under this view, a cryptic genetic variant may give rise to normally hidden phenotypes *when* it is altered in response to a stress. Such alterations, e.g., the result of stress-induced transposon insertions **(Fanti et al. 2017)** or IES retentions (this study), are at least partly nonrandom and might therefore reflect a preexisting adaptive developmental program in response to environmental adversities. This framework can account for why cryptic genetic variation may be preserved over evolutionary time. It predicts that the number of cryptic phenotypes in a population might differ based on the individuals' levels of genetic diversity or stress susceptibility. It also incorporates observations that indicate that stress-induced genetic changes are largely epigenetically controlled, may be trans-somatically inherited, and may reach fixation epigenetically (*e.g.*, **(Sollars et al. 2003)**). Much of

bioRxiv preprint doi: https://doi.org/10.1101/448316; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

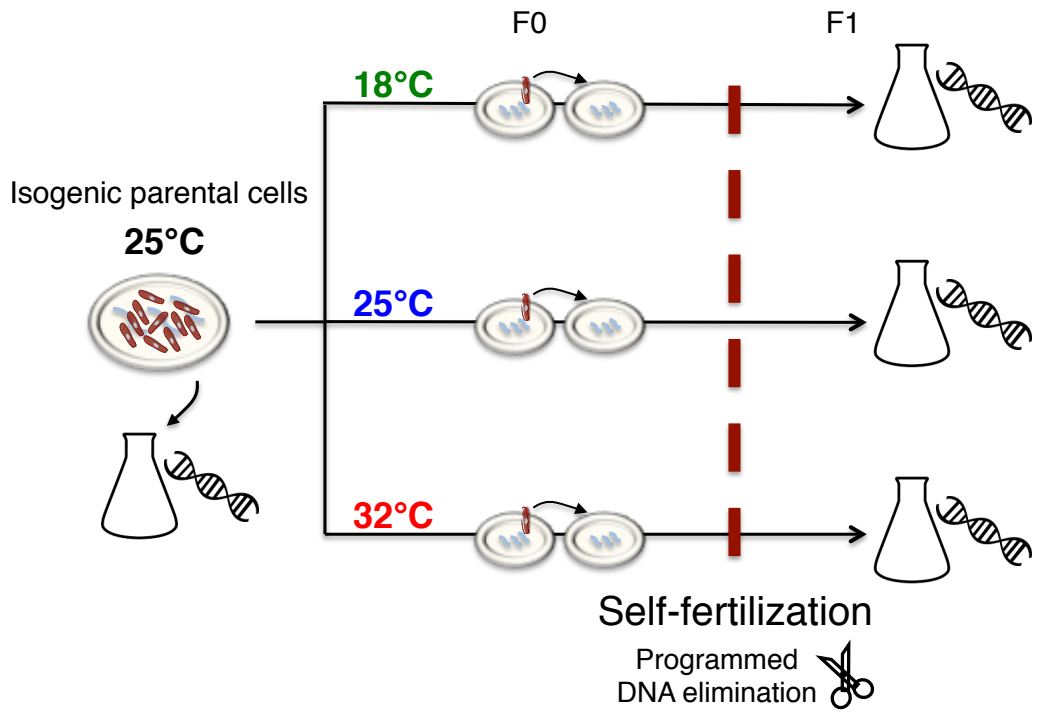
the stress-related genomic instability is likely to result from the disruption of small Piwi-interacting RNAs, which show several similarities between *Paramecium* and metazoans **(Bouhouche et al. 2011; Chalker and Yao 2011)**, and whose biogenesis is partially regulated by HSP90 **(Specchia et al. 2010; Gangaraju et al. 2011; Ichiyanagi et al. 2014)**.

In conclusion, we demonstrate that sub-optimal temperatures can modulate the efficiency of Programmed DNA Elimination in the ciliate *P. tetraurelia*, boosting the generation of epigenetically controlled, heritable, and functionally confined somatic DNA variability during germline-soma differentiation. The uncovered environmental sensitivity of Programmed DNA Elimination—a developmental process that unfolds in a broad range of unicellular and multicellular eukaryotes—is expected to elicit phenotypic plasticity in genetically identical organisms/cells. It also generates selectable variation, suggesting that Programmed DNA Elimination can operate as a molecular wrench that fine tunes organismal response to changing environmental conditions. Finally, our work reveals important similarities between *Paramecium* and multicellular organisms and, via these similarities, permits the elaboration of an adaptation model that is uniquely able to combine the previously argued adaptive roles of cryptic and *de novo* induced variation. Under this model, environmental cues affect the maternal environment, causing the epigenetic machinery that controls germline-soma differentiation to activate partially nonrandom plastic responses in the offspring generation. The resulting structural genomic changes are, at least partially, trans-somatically inherited and might facilitate adaptation to the triggering stress.

616 # Methods

617 ## Experimental Design

618 To evaluate the effect of the growth temperature on PDE, fully homozygous isogenic
619 *Paramecium* cells were cultured in daily re-isolation and passed through autogamy
620 (self-fertilization) at three different temperatures. After a first round of autogamy at
621 25˚C to establish a parental line, one post-autogamous parental cell was isolated and
622 allowed to divide in fresh medium (**Figure 8**, leftmost edge**)**. Three of the resulting
623 isogenic cells in turn were used to start three sub-lines that were cultured in daily re-
624 isolation and passed through a second round of autogamy at 18˚C, 25˚C, or 32˚C to
625 establish filial lines. Single post-autogamous F1 cells isolated for each of the sub-
626 lines as well as the remaining isogenic parental cells were expanded to mass culture
627 for somatic DNA extraction (see below).
628
629
630



631
632

633 **Figure 8**. Overview of the experimental setup to characterize the impact that temperature has on the
634 performance of Programmed DNA elimination in the ciliate *Paramecium tetraurelia* (see *Experimental*
635 *design*).
636

637 **Paramecium Strains and Culture conditions**
638 *Paramecium tetraurelia* strain d12 was used in the experiment. Cells were grown in
639 Cerophyl Medium (CM) inoculated with *Enterobacter aerogenes* Hormaeche and
640 Edwards (ATCC® 35028TM). Stocks were fed with bacterized CM every two weeks
641 and kept at 14˚C. In preparation for the study, cells kept in stock were washed ten
642 times in Volvic water to ensure monoxenic growth conditions during propagation.
643 Cells were grown in depression slides under daily re-isolation regimen in 200μl of
644 bacterized CM. Autogamy was induced by letting ~25 division-old cells starve
645 naturally for three days with no addition of bacterized medium. Occurrence of
646 autogamy was confirmed by screening ≥50 Acetocarmine-stained cells for
647 macronuclear fragmentation under a light microscope. A single ex-autogamous cell
648 was isolated from each line, allowed to divide once, and single caryonidal progenitors
649 brought to mass culture for somatic DNA isolation. Macronuclear fragmentation of
650 sister caryonides was confirmed by Acetocarmine staining.
651
652 **Macronuclear DNA Isolation and Whole Genome Sequencing**
653 The somatic nuclei of both parental (25˚C_{F0}) and filial lines (25˚C_{F1}, 32˚C_{F1} and
654 18˚C_{F1}) were isolated and the macronuclear DNA (MAC) subjected to whole genome
655 sequencing. MACs were isolated after >10 vegetative divisions post autogamy to
656 prevent carryover of maternal MAC fragments at the time of isolation. Cells were re-
657 suspended in Volvic water for 2 hours and allowed to digest their food vacuoles prior
658 to MAC isolation to reduce bacterial load. MAC isolation was performed according to
659 the protocol described in **(Arnaiz et al. 2012)**. Purified genomic DNA was subjected
660 to ultra-deep, pair-end Illumina sequencing (~90-100x coverage, on average, 150nt-
661 long reads) on a HiSeq 4000 system.
662
663 **Data preprocessing**
664 To increase the accuracy of DNA-seq data analysis raw reads were subjected to an
665 initial step of quality control (QC) using FastQC
666 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and pre-processed using
667 the Joint Genome Institute (JGI) suite BBTools 37.25 **(Bushnell et al. 2017)**. Library
668 adapters were removed from the 3' end of the reads with BBduk using the included
669 reference adapter file and ensuring the same length for both reads of a pair after
670 trimming. For the k-mer based adapter detection, a long and short k-mer size of 23
671 and 11, respectively, were used and a single mismatch allowed. In addition, overlap-
672 based adapter detection was enabled. Finally, short insert sizes in part of the library
673 were leveraged for an overlap-based error correction with BBMerge, while keeping
674 left and right reads separated.
675

676 **Calculation of IES retention scores**
677 Quality improved, trimmed reads were mapped to the *P. tetraurelia* strain 51
678 reference somatic genome available via ParameciumDB **(Arnaiz and Sperling 2011)**
679 and to a pseudo-germline genome containing all the known 44,928 IESs previously
680 identified by Piggy MAC (PGM) knock down **(Arnaiz et al. 2012)** that was created
681 with the Insert module of ParTIES 1.00 **(Denby Wilkes et al. 2016)**. Read mappings
682 were performed with Bowtie 2.3.2 **(Langmead et al. 2009)** using the local alignment
683 function for paired end reads in very sensitive mode (--very-sensitive-local) and the
684 resulting SAM files manipulated with SAMtools 1.4.1 **(Li et al. 2009)** for downstream
685 processing. IES retention scores (IRS) were calculated with the MIRET module of
686 ParTIES using the IES score method.
687
688 **Genome-wide analysis of IES retention**
689 Genome-wide analysis of IES retention was performed via statistical comparison of
690 F0/F1 IRSs as implemented in the R script accompanying ParTIES, with the following
691 modifications. The upper and lower bound of the 75% Confidence Interval (CI)
692 constructed on the F0 retention score was taken as a reference retention score for
693 binomial testing of upward or downward transitions, respectively. Lowly supported
694 IESs i.e. IESs with a total support < 20 reads (IES^+ + IES^-) were excluded from the
695 study. The *P*-values were corrected for multiple testing with the Benjamini-Hochberg
696 method and a cutoff of 0.05 used to designate IESs with significantly different
697 retention levels in F1 compared to F0 samples.
698
699 **Genome wide analysis of cryptic IES excisions**
700 To estimate the rate of cryptic IES excision in response to temperature changes
701 during nuclear differentiation TA-bound somatic deletions were characterized for the
702 parental F0 genome and the three F1 genomes rearranged at 18°C, 25°C and 32°C
703 using the MILORD module implemented in ParTIES. Reads mapped on the reference
704 macronuclear genome assembly of *P. tetraurelia* strain 51 were provided as input.
705 Low coverage cryptic IESs with total read counts (support_ref + support_variant) < 20
706 were excluded from the analysis. Deletion scores (DS) were calculated as the
707 fraction of reads supporting cryptic IES excision over all reads spanning the somatic
708 region, i.e. support_var/(support_var + support_ref). Cryptic IESs alternatively
709 excised across samples were collected by filtering deletion scores with standard
710 deviation > 0.3. Manual inspection with IGV **(Robinson et al. 2011)** was used to
711 confirm putative temperature-sensitive excisions that were tagged as 'Unstable'.
712 Conversely, cryptic IESs consistently excised (DS > 0.5 in all samples) were tagged
713 as 'Stable'. The catalog of cryptic IESs is provided in **Table S1**.

714
715 **Downstream Data Analyses**
716 We compiled a table (**Table S2,** provided as Supplementary Material) that reports the
717 obtained IRSs of all samples and a plethora of additional IES-related information
718 calculated *ad hoc* for this study or collected from previous studies and/or processed
719 from various external sources, *e.g.*, the *Paramecium tetraurelia* strain 51 genome
720 annotation v2 **(Arnaiz et al. 2017)**. All external data are publicly available at
721 ParameciumDB **(Arnaiz and Sperling 2011)**.The following downstream analyses
722 were conducted using in-house Python (https://www.python.org/) and R
723 (http://www.R-project.org) scripts.
724
725 **IES-Retention Profiles Simulation**
726 We simulated IES retention as a stochastic process. Four random samples were
727 drawn without replacement from the full reference set of IESs (PGM-IESs). Each
728 drawn sample contained as many elements as there were IESs with a non-zero
729 retention score detected for each of the four DNA samples. This simulation was
730 repeated 100 times and the maximum number of elements shared by all four
731 drawings (theoretical 4-way overlap assuming random IES retention) used as an
732 expectation to compare against the experimental data.
733
734 **Impact of Incomplete IES Excision on Genes**
735 For IESs located in the coding region of protein-coding genes we checked whether
736 their retention promotes the induction of a premature translation termination codon
737 (PTC). We calculated the IES's position with respect to the translation start codon,
738 inserting the IES at this location and scanning this artificial CDS+IES-construct for an
739 in-frame TGA (the only stop codon in *Paramecium*) upstream of the annotated one.
740 In case a PTC was detected, we marked the distance of the PTC to the CDS start
741 and, in case of non-3n IESs (IESs with a size that is not a multiple of 3), we marked
742 whether the PTC occurred inside the IES body or downstream.
743
744 **Gene families and GO-term enrichment analyses**
745 For each investigated temperature we tested whether specific gene families were
746 over or under-represented among the genes affected by significant IES retention. We
747 devised a sampling procedure that accounts for non-homogeneous IES densities
748 across *P. tetraurelia*' s genes. Briefly, 10000 IES sets with size equal to the number
749 of significantly retained IESs in the experimental samples were randomly drawn from
750 the PGM-IESs set. A non-redundant collection of genes hit by simulated IES
751 retention was extracted for each draw and the proportion of proteins carrying a
752 specific functional domain (e.g. TPR, Tetratrico Peptide Repeat-containing genes,

753 PKD, protein kinase-like domain) was used to build a null distribution. The mean of
754 this null distribution was then taken as the success probability for a two-sided
755 binomial test with a significance cutoff of $P < 0.01$. Raw *P*-values were adjusted via
756 the Benjamini–Hochberg procedure.
757
758 The functional enrichment analysis of IES-containing genes was performed using the
759 *statistical overrepresentation test* of the *Panther gene list analysis* tool **(Mi et al.**
760 **2013; Mi et al. 2017)** and cross-validated with the *topGO* package **(Alexa and**
761 **Rahnenfuhrer 2016)**. The mRNA IDs (e.g. GSPATT00000013001) of IES-containing
762 genes (macronuclear gene models v1) were used as supported gene identifiers for
763 the *Panther gene list analysis*. In all cases, IES-containing genes were tested against
764 the full set of macronuclear (coding) gene models of *P. tetraurelia*. The Weigh01
765 algorithm and F statistic (Fisher's exact test) were used for testing the GO-terms
766 overrepresentation with the *topGO* package. Raw *P*-values provided by *topGO* were
767 adjusted with the *P.adjust* function (method = "hochberg") implemented in the R
768 package *stats* (version 3.4.0). A critical value of 0.05 was adopted as significance
769 threshold in all tests.
770
771 **Data availability**
772 All the DNA sequence reads obtained for control and temperature exposed samples
773 have been deposited at the European Nucleotide Archive: PRJEB28697.

774

Acknowledgements

We thank Gennady Churakov, Franz Goller, and Hans-Dieter Görtz for their valuable comments on a draft of the manuscript. Kathrin Brüggemann is gratefully acknowledged for her technical assistance. This work was supported by a Deutsche Forschungsgemeinschaft (DFG) research grant to FC [CA1416/ 1-1] and carried out within the DFG Research Training Group 2220 'Evolutionary Processes in Adaptation and Disease' at the University of Münster.

Competing interests

The authors declare that no competing interests exist.

Author contributions

Valerio Vitali: *Investigation, Formal analysis, Software, Methodology, Visualization*, *Writing – original draft, Writing – review & editing.* Rebecca Hagen: *Data curation, Software, Writing – review & editing.* Francesco Catania: *Funding acquisition, Conceptualization, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.*

794 ## References

795 Alexa A, Rahnenfuhrer J. 2016. topGO: Enrichment Analysis for Gene Ontology. R
796 package version 2.32.0.
797 Allen SE, Nowacki M. 2017. Necessity Is the Mother of Invention: Ciliates,
798 Transposons, and Transgenerational Inheritance. *Trends Genet* **33**: 197-207.
799 Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury JM, Denby Wilkes C, Garnier O,
800 Labadie K, Lauderdale BE, Le Mouel A et al. 2012. The *Paramecium* Germline
801 Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics
802 of Internal Eliminated Sequences. *PLoS Genet* **8**: e1002984.
803 Arnaiz O, Sperling L. 2011. ParameciumDB in 2011: new tools and new data for
804 functional and comparative genomics of the model ciliate *Paramecium*
805 *tetraurelia*. *Nucleic Acids Res* **39**: D632-636.
806 Arnaiz O, Van Dijk E, Betermier M, Lhuillier-Akakpo M, de Vanssay A, Duharcourt S,
807 Sallet E, Gouzy J, Sperling L. 2017. Improved methods and resources for
808 paramecium genomics: transcription units, gene annotation and gene
809 expression. *BMC Genomics* **18**: 483.
810 Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Betermier M. 2009.
811 PiggyMac, a domesticated piggyBac transposase involved in programmed
812 genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* **23**:
813 2478-2483.
814 Betermier M, Duharcourt S. 2014. Programmed Rearrangement in Ciliates:
815 Paramecium. *Microbiol Spectr* **2**.
816 Biederman MK, Nelson MM, Asalone KC, Pedersen AL, Saldanha CJ, Bracht JR.
817 2018. Discovery of the First Germline-Restricted Gene by Subtractive
818 Transcriptomic Analysis in the Zebra Finch, *Taeniopygia guttata*. *Curr Biol* **28**:
819 1620-1627 e1625.
820 Bischerour J, Bhullar S, Denby Wilkes C, Regnier V, Mathy N, Dubois E, Singh A,
821 Swart E, Arnaiz O, Sperling L et al. 2018. Six domesticated PiggyBac
822 transposases together carry out programmed DNA elimination in *Paramecium*.
823 *Elife* **7**.
824 Bouhouche K, Gout JF, Kapusta A, Betermier M, Meyer E. 2011. Functional
825 specialization of Piwi proteins in *Paramecium tetraurelia* from post-
826 transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* **39**:
827 4249-4264.
828 Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B,
829 Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread
830 intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**:
831 1774-1786.
832 Brogna S, Wen J. 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat*
833 *Struct Mol Biol* **16**: 107-113.
834 Bushnell B, Rood J, Singer E. 2017. BBMerge - Accurate paired shotgun read
835 merging via overlap. *PLoS One* **12**: e0185056.
836 Caron F. 1992. A high degree of macronuclear chromosome polymorphism is
837 generated by variable DNA rearrangements in *Paramecium primaurelia* during
838 macronuclear differentiation. *J Mol Biol* **225**: 661-678.

839 Catania F, McGrath CL, Doak TG, Lynch M. 2013. Spliced DNA sequences in the
840 *Paramecium* germline: their properties and evolutionary potential. *Genome Biol*
841 *Evol* **5**: 1200-1211.
842 Catania F, Schmitz J. 2015. On the path to genetic novelties: insights from
843 programmed DNA elimination and RNA splicing. *Wiley Interdiscip Rev RNA* **6**:
844 547-561.
845 Chalker DL, Yao MC. 2011. DNA elimination in ciliates: transposon domestication and
846 genome surveillance. *Annu Rev Genet* **45**: 227-246.
847 Coyne RS, Lhuillier-Akakpo M, Duharcourt S. 2012. RNA-guided DNA rearrangements
848 in ciliates: is the best genome defence a good offence? *Biol Cell* **104**: 309-325.
849 Denby Wilkes C, Arnaiz O, Sperling L. 2016. ParTIES: a toolbox for *Paramecium*
850 interspersed DNA elimination studies. *Bioinformatics* **32**: 599-601.
851 Dubois E, Bischerour J, Marmignon A, Mathy N, Regnier V, Betermier M. 2012.
852 Transposon Invasion of the Paramecium Germline Genome Countered by a
853 Domesticated PiggyBac Transposase and the NHEJ Pathway. *International*
854 *journal of evolutionary biology* **2012**: ID 436196.
855 Duharcourt S, Keller AM, Meyer E. 1998. Homology-dependent maternal inhibition of
856 developmental excision of internal eliminated sequences in *Paramecium*
857 *tetraurelia*. *Mol Cell Biol* **18**: 7075-7085.
858 Duharcourt S, Lepere G, Meyer E. 2009. Developmental genome rearrangements in
859 ciliates: a natural genomic subtraction mediated by non-coding transcripts.
860 *Trends Genet* **25**: 344-350.
861 Duret L, Cohen J, Jubin C, Dessen P, Gout JF, Mousset S, Aury JM, Jaillon O, Noel B,
862 Arnaiz O et al. 2008. Analysis of sequence variability in the macronuclear DNA
863 of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res* **18**:
864 585-596.
865 Fanti L, Piacentini L, Cappucci U, Casale AM, Pimpinelli S. 2017. Canalization by
866 Selection of de Novo Induced Mutations. *Genetics* **206**: 1995-2006.
867 Ferro D, Lepennetier G, Catania F. 2015. *Cis*-acting signals modulate the efficiency of
868 programmed DNA elimination in *Paramecium tetraurelia*. *Nucleic Acids Res* **43**:
869 8157-8168.
870 Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002.
871 An mRNA surveillance mechanism that eliminates transcripts lacking
872 termination codons. *Science* **295**: 2258-2261.
873 Gangaraju VK, Yin H, Weiner MM, Wang J, Huang XA, Lin H. 2011. Drosophila Piwi
874 functions in Hsp90-mediated suppression of phenotypic variation. *Nat Genet*
875 **43**: 153-158.
876 Garnier O, Serrano V, Duharcourt S, Meyer E. 2004. RNA-mediated programming of
877 developmental genome rearrangements in *Paramecium tetraurelia*. *Mol Cell*
878 *Biol* **24**: 7370-7379.
879 Hoehener C, Hug I, Nowacki M. 2018. Dicer-like Enzymes with Sequence Cleavage
880 Preferences. *Cell* **173**: 234-247 e237.
881 Ichiyanagi T, Ichiyanagi K, Ogawa A, Kuramochi-Miyagawa S, Nakano T, Chuma S,
882 Sasaki H, Udono H. 2014. HSP90alpha plays an important role in piRNA
883 biogenesis and retrotransposon repression in mouse. *Nucleic Acids Res* **42**:
884 11903-11911.

885 Jarosz DF, Lindquist S. 2010. Hsp90 and environmental stress transform the adaptive
886 value of natural genetic variation. *Science* **330**: 1820-1824.
887 Jollos V. 1921. *Experimentelle Protistenstudien: Untersuchungen über Variabilität und*
888 *Vererbung bei Infusorien. (Experimental Studies on Protists: Studies on*
889 *Variability and Inheritance in Infusoria)*. G. Fischer, Jena
890 Jollos V. 1934. Inherited changes produced by heat-treatment in *Drosophila*
891 *melanogaster*. *Genetica* **16**: 476-494.
892 Jung D, Giallourakis C, Mostoslavsky R, Alt FW. 2006. Mechanism and control of
893 V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev*
894 *Immunol* **24**: 541-570.
895 Kasinathan B, Ahmad K, Malik HS. 2017. Waddington Redux: De Novo Mutations
896 Underlie the Genetic Assimilation of Stress-Induced Phenocopies in *Drosophila*
897 *melanogaster*. *Genetics* **207**: 49-51.
898 Klauer AA, van Hoof A. 2012. Degradation of mRNAs that lack a stop codon: a
899 decade of nonstop progress. *Wiley Interdiscip Rev RNA* **3**: 649-660.
900 Klobutcher LA, Herrick G. 1995. Consensus inverted terminal repeat sequence of
901 *Paramecium* IESs: resemblance to termini of Tc1-related and Euplotes Tec
902 transposons. *Nucleic Acids Res* **23**: 2006-2013.
903 Krenek S, Berendonk TU, Petzoldt T. 2011. Thermal performance curves of
904 Paramecium caudatum: a model selection approach. *Eur J Protistol* **47**: 124-
905 137.
906 Krenek S, Petzoldt T, Berendonk TU. 2012. Coping with temperature at the warm
907 edge--patterns of thermal adaptation in the microbial eukaryote Paramecium
908 caudatum. *PLoS One* **7**: e30598.
909 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
910 alignment of short DNA sequences to the human genome. *Genome Biol* **10**:
911 R25.
912 Lea AJ, Tung J, Archie EA, Alberts SC. 2017a. Developmental plasticity research in
913 evolution and human health: Response to commentaries. *Evol Med Public*
914 *Health* **2017**: 201-205.
915 Lea AJ, Tung J, Archie EA, Alberts SC. 2017b. Developmental plasticity: Bridging
916 research in evolution and human health. *Evol Med Public Health* **2017**: 162-
917 175.
918 Lepere G, Betermier M, Meyer E, Duharcourt S. 2008. Maternal noncoding transcripts
919 antagonize the targeting of DNA elimination by scanRNAs in *Paramecium*
920 *tetraurelia*. *Genes Dev* **22**: 1501-1512.
921 Lepere G, Nowacki M, Serrano V, Gout JF, Guglielmi G, Duharcourt S, Meyer E.
922 2009. Silencing-associated and meiosis-specific small RNA pathways in
923 *Paramecium tetraurelia*. *Nucleic Acids Res* **37**: 903-915.
924 Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of
925 alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl*
926 *Acad Sci U S A* **100**: 189-192.
927 Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L,
928 Duharcourt S. 2014. Local Effect of Enhancer of Zeste-Like Reveals
929 Cooperation of Epigenetic and cis-Acting Determinants for Zygotic Genome
930 Rearrangements. *PLoS Genet* **10**: e1004665.

931 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
932 Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and
933 SAMtools. *Bioinformatics* **25**: 2078-2079.
934 Lloyd A, Morgan C, FC HF, Bomblies K. 2018. Plasticity of Meiotic Recombination
935 Rates in Response to Temperature in Arabidopsis. *Genetics* **208**: 1409-1420.
936 Marquez Y, Hopfler M, Ayatollahi Z, Barta A, Kalyna M. 2015. Unmasking alternative
937 splicing inside protein-coding exons defines exitrons and their role in proteome
938 plasticity. *Genome Res* **25**: 995-1007.
939 Mayer KM, Forney JD. 1999. A mutation in the flanking 5'-TA-3' dinucleotide prevents
940 excision of an internal eliminated sequence from the *Paramecium tetraurelia*
941 genome. *Genetics* **151**: 597-604.
942 Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017.
943 PANTHER version 11: expanded annotation data from Gene Ontology and
944 Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*
945 **45**: D183-D189.
946 Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function
947 analysis with the PANTHER classification system. *Nat Protoc* **8**: 1551-1566.
948 Noto T, Mochizuki K. 2017. Whats, hows and whys of programmed DNA elimination in
949 Tetrahymena. *Open Biol* **7**.
950 Noto T, Mochizuki K. 2018. Small RNA-Mediated trans-Nuclear and trans-Element
951 Communications in Tetrahymena DNA Elimination. *Curr Biol* **28**: 1938-1949
952 e1935.
953 Orias E, Singh DP, Meyer E. 2017. Genetics and Epigenetics of Mating Type
954 Determination in Paramecium and Tetrahymena. *Annu Rev Microbiol* **71**: 133-
955 156.
956 Prescott DM. 1994. The DNA of ciliated protozoa. *Microbiol Rev* **58**: 233-267.
957 Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic
958 variation. *Nature* **417**: 618-624.
959 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov
960 JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.
961 Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, Lindquist
962 S, Tabin CJ. 2013. Cryptic variation in morphological evolution: HSP90 as a
963 capacitor for loss of eyes in cavefish. *Science* **342**: 1372-1375.
964 Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution.
965 *Nature* **396**: 336-342.
966 Sandoval PY, Swart EC, Arambasic M, Nowacki M. 2014. Functional diversification of
967 Dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell* **28**:
968 174-188.
969 Singh DP, Saudemont B, Guglielmi G, Arnaiz O, Gout JF, Prajer M, Potekhin A,
970 Przybos E, Aubusson-Fleury A, Bhullar S et al. 2014. Genome-defence small
971 RNAs exapted for epigenetic mating-type inheritance. *Nature* **509**: 447-452.
972 Singh P, Borger C, More H, Sturmbauer C. 2017. The Role of Alternative Splicing and
973 Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biol Evol*
974 **9**: 2764-2781.

975 Smith CCR, Tittes S, Mendieta JP, Collier-Zans E, Rowe HC, Rieseberg LH, Kane
976 NC. 2018a. Genetics of alternative splicing evolution during sunflower
977 domestication. *Proc Natl Acad Sci U S A* **115**: 6768-6773.
978 Smith JJ, Timoshevskaya N, Ye C, Holt C, Keinath MC, Parker HJ, Cook ME, Hess
979 JE, Narum SR, Lamanna F et al. 2018b. The sea lamprey germline genome
980 provides insights into programmed genome rearrangement and vertebrate
981 evolution. *Nat Genet* **50**: 270-277.
982 Sollars V, Lu X, Xiao L, Wang X, Garfinkel MD, Ruden DM. 2003. Evidence for an
983 epigenetic mechanism by which Hsp90 acts as a capacitor for morphological
984 evolution. *Nat Genet* **33**: 70-74.
985 Sonneborn TM. 1947. Recent Advances in the Genetics of Paramecium and Euplotes.
986 In *Advances in Genetics*, Vol 1 (ed. M Demerec), pp. 263-358. Academic
987 Press.
988 Sonneborn TM. 1977. Genetics of cellular differentiation: stable nuclear differentiation
989 in eucaryotic unicells. *Annu Rev Genet* **11**: 349-367.
990 Sonneborn TM, M.V. S. 1979. A Genetic System for Alternative Stable Characteristics
991 in Genomically Identical Homozygous Clones. *Developmental Genetics* **1**: 21–
992 46.
993 Specchia V, Piacentini L, Tritto P, Fanti L, D'Alessandro R, Palumbo G, Pimpinelli S,
994 Bozzetti MP. 2010. Hsp90 prevents phenotypic variation by suppressing the
995 mutagenic activity of transposons. *Nature* **463**: 662-665.
996 Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. 2018. Developmental
997 Bias and Evolution: A Regulatory Network Perspective. *Genetics* **209**: 949-966.
998 Vasudevan S, Peltz SW, Wilusz CJ. 2002. Non-stop decay--a new mRNA surveillance
999 pathway. *Bioessays* **24**: 785-788.
1000 Waddington CH. 1953. Genetic assimilation of an acquired character. *Evolution* **7**:
1001 118-126.
1002 Wang J, Davis RE. 2014. Programmed DNA elimination in multicellular organisms.
1003 *Curr Opin Genet Dev* **27**: 26-34.
1004 West-Eberhard MJ. 2005. Developmental plasticity and the origin of species
1005 differences. *Proc Natl Acad Sci U S A* **102 Suppl 1**: 6543-6549.
1006 Wong JJ, Ritchie W, Ebner OA, Selbach M, Wong JW, Huang Y, Gao D, Pinello N,
1007 Gonzalez M, Baidya K et al. 2013. Orchestrated intron retention regulates
1008 normal granulocyte differentiation. *Cell* **154**: 583-595.
1009 Yeyati PL, Bancewicz RM, Maule J, van Heyningen V. 2007. Hsp90 selectively
1010 modulates phenotype in vertebrate development. *PLoS Genet* **3**: e43.

1011