The Pathway to Detangle a Scrambled Gene

Matthias Möllenbeck1, Yi Zhou2, Andre R. O. Cavalcanti2a, Franziska Jönsson1, Brian P. Higgins2, Wei-Jen Chang2b, Stefan Juranek1, Thomas G. Doak2, Grzegorz Rozenberg3, Hans J. Lipps1, Laura F. Landweber2

1 Institute of Cell Biology, University Witten/Herdecke, Witten, Germany, 2 Ecology & Evolutionary Biology, Princeton University, Princeton, New Jersey, United States of America, 3 Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

Abstract

Background: Programmed DNA elimination and reorganization frequently occur during cellular differentiation. Development of the somatic macronucleus in some ciliates presents an extreme case, involving excision of internal eliminated sequences (IESs) that interrupt coding DNA segments (macronuclear destined sequences, MDSs), as well as removal of transposon-like elements and extensive genome fragmentation, leading to 98% genome reduction in Styloynchia lemmatae. Approximately 20–30% of the genes are estimated to be scrambled in the germline micronucleus, with coding segment order permuted and present in either orientation on micronuclear chromosomes. Massive genome rearrangements are therefore critical for development.

Methodology/Principal Findings: To understand the process of DNA deletion and reorganization during macronuclear development, we examined the population of DNA molecules during assembly of different scrambled genes in two related organisms in a developmental time-course by PCR. The data suggest that removal of conventional IESs usually occurs first, accompanied by a surprising level of error at this step. The complex events of inversion and translocation seem to occur after repair and excision of all conventional IESs and via multiple pathways.

Conclusions/Significance: This study reveals a temporal order of DNA rearrangements during the processing of a scrambled gene, with simpler events usually preceding more complex ones. The surprising observation of a hidden layer of errors, absent from the mature macronucleus but present during development, also underscores the need for repair or screening of incorrectly-assembled DNA molecules.

Introduction

DNA elimination and reorganization occur in a variety of differentiating eukaryotic cells. Genome reduction may involve either whole chromosomes, as in differentiating cells of Sciara coprophila [1] or parts of chromosomes, as in nematodes or Cyclops [2–4]. The best studied example of specific DNA excision and rearrangement is the processing of immunoglobulin and T-cell receptor genes in mammalian cells [5,6]. The mechanisms of all these rearrangements are not completely understood, but most cases probably involve recombination [7,8], similar to mating type switching in yeast [9,10] or antigenic variation in trypanosomes [11,12]. The most elaborate known form of DNA rearrangement occurs during macronuclear development in ciliated protists, which provide model systems to study programmed DNA elimination, fragmentation, and reorganization during development [13–18].

Most ciliates possess two types of nuclei: a transcriptionally active macronucleus (soma) and a germline micronucleus used for sexual conjugation. After sexual reproduction, the diploid zygotic micronucleus develops into a DNA-rich macronucleus, with each chromosome amplified to a high copy number, with some variation. Non-coding DNA segments (IESs) [19] interrupt most of the ~30,000 genes in the germline genome of stichotrichous ciliates. Macronuclear development in these species removes intergenic DNA as well as deleting more than 100,000 IESs to allow assembly of the gene segments (MDSs) into functional genes. Fully-assembled genes typically reside on minimalist “gene-sized nanochromosomes” with short telomeres on both ends and almost no non-coding DNA. More surprisingly, the process of gene assembly can even invert and/or permute (switch) segment order, and scrambled genes requiring these complicated events may account for 20–30% of all genes in stichotrichs. For example, actin I, telomere-end-binding protein subunit α, and DNA polymerase α, among other genes (e.g. [15,16,20–22]), contain their germline segments in a permuted order relative to their orthodox (unscrambled) order in the macronucleus. Scrambled segments must be reordered and...
IESs removed to construct functional, translatable genes in the macronucleus [18].

Homologous recombination between short sequence repeats, called pointers, at MDS–IES junctions is involved in gene unscrambling, facilitating both removal of IESs and sorting of MDSs [29]. For example, a DNA sequence present at the junction between MDS n and the downstream IES is generally identical to a sequence between MDS n+1 and its upstream IES, leading to correct fusion of segment n to n+1, even over long distances. Despite their presence at all known MDS junctions, pointer repeats are short (average repeat length 4 bp between non-scrambled segments, 9 bp between scrambled segments [24]; current data still agree with these estimates) sometimes with mismatches, implying that pointer recognition alone is insufficient to direct accurate splicing. The repeats may satisfy a structural requirement for DNA splicing, and less of a role in recognition; otherwise incorrectly spliced sequences (results of promiscuous recombination) would dominate. While illegitimate recombination might be a source of new scrambled patterns in the germline [25], macronuclear development retains only the molecules that acquire telomerases at both ends [26], possibly ensuring loss of some promiscuously ordered genes.

The molecular mechanisms of gene unscrambling are not yet understood. Recent studies in several species of ciliates indicate involvement of an RNAi-related mechanism to direct genomewide DNA rearrangements that may have arisen as a defense against invading genetic agents [27]. A model introduced for Tetrahymena and extended to Paramecium proposes that small RNAs might function to tag sequences for elimination by a mechanism similar to RNA-mediated gene silencing [28–30] (reviewed in [31]). A limitation of this model for stichotrichs is the complexity of events in gene unscrambling. Furthermore, the IESs in these species may be too small to create a specific chromatin structure [32]. Because gene unscrambling requires such an amazing level of precision, current models [29,33–36] with experimental evidence [36] propose that a specific set of templates derived from the old macronucleus may guide assembly, simultaneously repairing imprecise IES excision and directing MDS sorting [33].

To begin to unravel the events of DNA excision and unscrambling during development, we investigated the order of DNA elimination, inversion and reordering in the actin I gene in Stylonychia lemnae, in a developmental time-course. We chose this locus as our main model system, because it is the simplest scrambled gene whose assembly requires all three types of DNA rearrangements, including deletion, inversion, and permutation, or segment reordering [37].

Structure of the scrambled genes

The structures of the micronuclear genes for actin I in Stylonychia lemnae and its ortholog in Oxytricha trifallax, as well as O. trifallax TEBPs are shown in Figure 1. The S. lemnae actin I gene contains 10 MDSs in the scrambled germline order: 3 1 4 5 6 7 8*10*2-1*9, with segments 1 and 2 inverted. Of the nine IESs, six are conventionally spliced and three, indicated by asterisks, define scrambled junctions. Assembly of actin I in S. lemnae therefore requires at least six events of conventional DNA deletion (joining segments 3–8 and 1–2), plus insertion of segment 9 between segments 8 and 10 and inversion of segments 1 and 2, as well as their translocation to the 5’ end of the molecule (or inversion of segments 3–10) (see Results section on DNA Permutation for more details). Note that the joining of segments 2 to 3, 8 to 9, and 9 to 10 all require permutations. The repeats (pointers) at the boundaries between MDS and IES in these scrambled genes range from 3–19 bp (Table 1–3). In [37] we examined the evolution of the actin I gene structure in two geographically isolated strains of S. lemnae, which differ significantly in non-coding regions. This study uses only the German strain.

Results

Time-course of DNA deletion during macronuclear development

To understand the process of IES excision and DNA descrambling, we first examined the population of molecules present during assembly of the actin I gene in S. lemnae. We PCR amplified DNA from different stages of macronuclear development (called Anlage, Figure 2). Anlagen DNA was isolated from partially-synchronized exconjugants, which yield some variation. Five time-points covered a 30-hour period from early polytene

Figure 1. Schematic representation of the micronuclear precursor and resulting macronuclear gene structures (to scale) for (A) Stylonychia lemnae (D strain) actin I, (B) Oxytricha trifallax actin I, and (C) Oxytricha trifallax TEBPs. MDSs represented in blue and labeled m (upside down if inverted); IESs labeled i and flanking DNA in yellow; pointer sequences, when shown, in maroon. Triangles mark the positions of start (pointing down) and stop codons (pointing up). Black boxes on both ends of the macronuclear structure represent telomerases.

doi:10.1371/journal.pone.0002330.g001
chromosome formation to the beginning of DNA degradation (Figure 2). We amplified and sequenced a broad ensemble of partially-rearranged molecules at each time interval using multiple sets of PCR primers (primer sequences are listed in Methods S1; positions are shown in Figure 3 and Figure S1). We then used these “snapshots” of the range of partially-processed molecules to infer the time-course of rearrangement. In addition, we extended this approach to two genes in a related ciliate, *O. trifallax*, to help document the authenticity of partially-rearranged molecules and to avoid being misled by one gene-specific study. The time points used for *O. trifallax* cover a wider range of development stages, from 0-hour (just after mixing of mating types) to 55-hours (DNA poor stage).

To avoid contamination from maternal macronuclear DNA that has not been fully degraded, the sequence and/or orientation of most *S. lemnae* primer-pairs (Methods S1 and Figure S1) cannot amplify the fully-rearranged product, although they can amplify the unprocessed or partially processed micronuclear precursor (Table S1). This strategy enabled us to focus on partially-processed molecules that are incompletely rearranged. For example, the combination of a forward primer from the end of segment 3 (before the first IES) with a reverse primer from inverted segments 2 and/or 1 avoided macronuclear DNA contamination and permitted dense sampling of partially-processed molecules that lacked inversions. In the few cases in which we used primers that could amplify both micronuclear and macronuclear segment order, we recovered similar types of partially-processed molecules, suggesting that most of the observed DNA rearrangements are not artifacts of PCR or primer amplification bias. Furthermore, recent RNAi experiments (See Figures 1b and d in [36]) that disrupted the process of gene unscrambling in *O. trifallax* led to an accumulation of similar partially-processed molecules for two different scrambled genes (*TEBP* and *DNA polymerase*), providing independent support for the authenticity of incorrect or partially-rearranged molecules. While this dataset [36] was also generated by PCR, the incorrectly-processed molecules were in greater abundance (permitting PCR cycle number to be small), because the process of rearrangement had been experimentally stalled or halted by depleting specific RNA templates for two independent genes via RNAi. For greater resolution in the present experiments, particularly of scrambled and inverted regions, we amplified shorter intervals, because of the difficulty of capturing

| Table 1. Characteristics of MDs, IESs, and pointers of the *S. lemnae* actin I gene. |
| MDS m | Length* of MDS m | Pointer sequence between MDS m and m-1 | IES i | between MDs | Length* of IES i |
|-------|-----------------|----------------------------------------|-------|-------------|-----------------|
| 1     | 85              | CACAT                                  | 8     | 1–2         | 64              |
| 2     | 82              | GAGT(A/C)GCTCAAGGCTGGTTTT              | 7b    | 10–2        | 15              |
| 3     | 590             | TCGTT                                  | 1     | 3–4         | 23              |
| 4     | 149             | AAGA                                   | 2     | 4–5         | 49              |
| 5     | 117             | ATT                                    | 3     | 5–6         | 46              |
| 6     | 192             | TGC                                    | 4     | 6–7         | 29              |
| 7     | 27              | GTTAATTAT                               | 5     | 7–8         | 39              |
| 8     | 28              | GAATCA                                  | 6     | 8–10        | 17              |
| 9     | 43              | GCCAAGGACAGGGTGA                       | 9b    | 1–9         | 29              |
| 10    | 86              |                                        |       |             |                 |

Inverted MDS segments are represented by a bar. MDS numbers *m* are consecutive in the macronucleus, and IES numbers *i* are consecutive in the micronucleus, as in Figure 1. The scrambled micronuclear pointers that link non-consecutive MDSs are boldfaced.

*Lengths of MDs and IESs exclude pointers.
*These IESs contain a telomere addition site on one side (beginning of MDS 1 and end of MDS 10) and a scrambled pointer on the other side.

doi:10.1371/journal.pone.0002330.t001

| Table 2. Characteristics of MDs, IESs, and pointers of the *O. trifallax* actin I gene. |
| MDS m | Length* of MDS m | Pointer sequence between MDS m and m-1 | IES i | between MDs | Length* of IES i |
|-------|-----------------|----------------------------------------|-------|-------------|-----------------|
| 1     | 190             | GACCAACAAA                             | 9     | 1–8         | 28              |
| 2     | 32              | AAGGCCGTTGTTTGC                        | 8a    | 2–1         | 26              |
| 3     | 227             | TCTC                                   | 1     | 3–4         | 69              |
| 4     | 566             | AGCTCCCAAAGTCAA                        | 2     | 4–6         | 29              |
| 5     | 12              | TTATTCGCCA                             | 4     | 5–7         | 21              |
| 6     | 31              | TGAGGAAT                               | 3     | 6–5         | 53              |
| 7     | 75              | TGATACCTAAC                            | 5     | 7–9         | 19              |
| 8     | 17              | GGGTGGATGTA                            |       |             |                 |
| 9     | 50              | CAAAAAT                                | 6     | 9–10        | 215             |
| 10    | 17              |                                        | 7a    | 10–2        | 22              |

See Table 1 for notation.
doi:10.1371/journal.pone.0002330.t002
long, partially-processed molecules by PCR. We take this difficulty in amplification to indicate either rarity or transience of such molecules, or the presence of nicked or heavily supercoiled DNA [38].

Figure 3 summarizes all conventional DNA deletion events involving non-scrambled IESs 1–5 and 8 in S. lemnae actin I. IESs are still present in the scrambled regions. All conventional IESs are deleted from most surveyed molecules by the middle-late to late stage of polytene chromosome formation. However, unlike a previous study that used PCR and Southern hybridization to track processing of two simple loci containing just two and three nonscrambled IESs, respectively [39], we detected no specific order of conventional DNA deletion events.

Surprisingly, we found that conventional IES excision can actually be very imprecise, with a high level of error at MDS junctions among the molecules surveyed during IES removal (Table S2 and Figure S2). Although the use of primer combinations that amplify the micronuclear-specific segment order may favor molecules that contain errors or are potentially developmentally “stalled”, especially during the middle-late to late stage, we detected similar errors in partially-processed molecules amplified with primers that can amplify both micronuclear and macronuclear segment order, again suggesting that these errors are not experimental artifacts. We also recovered many of the same types of excision errors multiple times from independent experiments and different primers, indicating that the data are unlikely to be PCR artifacts. The data for two genes in O. trifallax (Figure S3) contain the same types of partially-processed errors.

All imprecisely joined MDSs in both species were fused at 1–8 nucleotide direct repeats, or cryptic pointers, flanking the excised sequence in the precursor molecule, and usually in close proximity to, or within the authentic pointers (details in Figure S2A). The use of cryptic pointers is consistent with a mechanism involving recombination at regions of micro-homology, possibly pausing at such repeats. Errors of both excessive deletion (erosion of coding segments) and insufficient deletion (retention of complete or partial IESs) are present. Assuming that the accuracy of excision at different boundaries is independent, and that the error rates we detect are not highly biased, we can roughly estimate (Table S3) that most molecules are likely to contain at least one incorrect deletion event during development. Since most of these incorrect deletion events disrupt reading frame or protein sequence, in the absence of a selection step or polymerase-mediated proofreading, we can expect that more than three quarters of partially-processed molecules might yield dead-end products, rather than developing into functional macronuclear chromosomes.

Our data (Table S3) do not suggest any correlation between the accuracy of conventional IES deletion and pointer or IES length; i.e. the processing of IESs with longer pointers or shorter length is no more robust. However, among cryptic pointers, longer repeats are used more often in spurious rearrangements (Figure S2A and Table S2A). These observations suggest that the events of IES removal may favor longer repeats, although the final accuracy of

| MDS m | Length* of MDS m | Pointer sequence between MDS m and m+1 | IES i | between MDSs | Length* of IES i |
|-------|-----------------|--------------------------------------|-------|---------------|----------------|
| 1     | 196             | TAC(A/G)AAGAA                       | 1     | 1–3           | 42             |
| 2     | 293             | TTCTCTTG                           | 7     | 2–4           | 38             |
| 3     | 41              | CACAGACTTGGAG                      | 2     | 3–5           | 6              |
| 4     | 21              | AAGAT(T/C)CACA                      | 8     | 4–6           | 22             |
| 5     | 160             | A(A/C)TCAG                         | 3     | 5–7           | 32             |
| 6     | 33              | AGAAGAAGA                         | 9     | 6–8           | 4              |
| 7     | 18              | TGTCTAAAAC                        | 4     | 7–10          | 26             |
| 8     | 30              | AAT                                | 10    | 8–9           | 46             |
| 9     | 138             | AGCTTAAG                           | 11    | 9–11          | 20             |
| 10    | 19              | TCAAGTTTTT(C/T)T                   | 5     | 10–12         | 9              |
| 11    | 56              | GGTG(T/C)C                        | 12    | 11–13         | 14             |
| 12    | 17              | TCAGCTACTTA                       | 6     | 12–2          | 52             |
| 13    | 28              | AAAAGAA                            | 13    | 13–14         | 40             |
| 14    | 165             | ACAAGAA                           | 14    | 14–15         | 64             |
| 15    | 239             | TCGTTA                            | 15    | 15–16         | 51             |
| 16    | 235             | CAGATT                             | 16    | 16–17         | 107            |
| 17    | 397             |                                     |       |               |                |

See Table 1 for notation.

doi:10.1371/journal.pone.0002330.t003

Figure 2. Macronuclear development in Stylonychia lemnae (adopted from [49,52] and consistent with present lab conditions) showing time-points for DNA isolation (e= early; e-m= early-middle; m= middle; m-l= middle-late; and l= late).

doi:10.1371/journal.pone.0002330.g002
excision may depend on a step that specifically degrades incorrect products or uses template-directed proofreading [29,33–36].

DNA Permutation

To address segment reordering, we amplified DNA from different stages of development (Figure 2) using four types of actin I primer combinations in S. lemnae (Figure S1): (A) forward and reverse primers derived from the 5' and 3' flanking regions; (B) a forward primer at the non-scrambled 5' end of the germline gene and a reverse primer in the 3' flanking region; (C) a forward primer in the 5' flanking region and a reverse primer at the 3' end of the germline gene; and (D) a pair of primers within the micronuclear gene. (A) and (B) yielded products with only conventional IESs removed but no permutations. On the other hand, (C) yielded two types of molecules: The first group (marked * in Figure 3) deleted some conventional IESs, with no segments unscrambled. The second group (Mn101005-20 and Mn101005-14 in Figure S5) deleted ALL conventional IESs and correctly relocated segment 9. Combination (D) produced both unpermuted and permuted rearrangements. The presence of PCR products from (A) with conventional IESs removed suggests that DNA deletion may precede chromosome fragmentation and rearrangement. The lack of rearranged PCR products in combination (B) suggests that fragmentation occurs first at or near the 3' end (segment 9) prior to descrambling segment 9 (or any other permutations). Breakage at the 5' end could occur after segment 9 translocation, as supported by the presence of PCR products from (C) that contain both 5' flanking DNA and segment 9 unscrambled.

Assembly of a functional S. lemnae actin I micronuclear gene requires inversion of segments 1 and 2, coupled with translocation to the 5' end (or inversion of segments 3–10), plus insertion of segment 9 between 8 and 10, yielding two possible pathways for the permutation events during actin I descrambling (Figure 4). We recovered two classes of partially-rearranged molecules, including potential unscrambling intermediates that support both possible
pathways (Figure 4). (Some aberrantly rearranged products were also detected that we discuss below.) In the first class of molecules, segment 9 was correctly positioned between 8 and 10, with segments 2 and 1 still in their original inverted orientation at the 3’ end of the molecule. In the second class, either segments 2 and 1 together with segment 9 may be inverted and linked to segment 3 at the 5’ end, or segments 3–10 are inverted and joined to segment 1–2; however, these molecules did not survey processing at the 3’ end, and hence we do not know whether segment 9 is also present between 8 and 10. In addition, three of the four molecules supporting the second pathway also contained cryptic deletions, consistent with the types of errors observed elsewhere. More putative intermediates of the second type were recovered for O. trifallax actin I [Figure S4A cases a–c] in an unbiased PCR experiment (primers located in segments 1 and 9) that covered both ends of the molecule, supporting the authenticity of this possible pathway. No examples in O. trifallax supported the first pathway (Figure S3A and S4A); however, but this could be due to small sample size. Therefore, both actin I descrambling pathways may be evolutionarily feasible, with species-specific biases possible for one or the other pathway.

Some molecules with segment 9 translocations contain another copy of segment 10 at its original micronuclear position (Mn101005-20 and Mn101005-14 in Figure S5), suggesting the possibility of either intermolecular recombination or additional error; however we cannot exclude intra-molecular replication or PCR error to explain the segment 10 duplications. The data for TEBPα (Figure S4B) suggest that joining of segments 2–3, 5–4, or 4–5 may occur independently, suggesting multiple pathways for unscrambling this complex gene.

We also detected various aberrantly rearranged products for S. lemnæ actin I that do not fit into a possible unscrambling pathway (Figure 5). All but one molecule (case g) contain repeats of 1–10 bp as possible cryptic pointers at their rearranged junctions (details in Table S2B). Aberrant rearrangements can affect both scrambled and non-scrambled regions, displacing pointers at conventional IESs as well as scrambled junctions. Although the cryptic pointers at aberrant junctions are often longer than the non-scrambled pointers they replace (Table S2), they are generally in the same length range as scrambled pointers and usually not longer than the original, displaced scrambled pointer. This suggests that factors other than repeat length drive rearrangement events. Other influences may include the physical proximity of DNA segments in the developing macronucleus. Consistent with this hypothesis, we detect no correlation between the accuracy or robustness of a scrambled junction and the length of its correct pointers (Table S4); however, the junctions between segments 8 and 9 and segments 9 and 10, which are closer to each other in the precursor sequence than the junctions between segments 2 and 3, are much more robust.

Intriguingly, in S. lemnæ we only recovered molecules with permutations (Figure 5) or large deletions (lines marked Δ in Figure 3; detail in Figure S2B) in the middle-late or late stage of polytene chromosome formation. With just two exceptions (cases b and h in Figure 5), all molecules containing permutations (both correct and aberrant) lack IES sequences entirely, and furthermore contain only correctly excised (or correctly healed) junctions at all nonscrambled IESs. The O. trifallax actin I and TEBPα data are similar (Fig. S4), although more partially reordered actin I molecules in O. trifallax retain a small conventional IES. This suggests that the cell executes gross rearrangements, usually by inversion and/or translocation, often after excision and possible repair of all conventional IES junctions, indicating the presence of preferred steps in the gene assembly process. These steps could be associated with the expression onset of specific proteins or factors responsible for different types of reactions during DNA rearrange-
Discussion

We surveyed the population of partially processed DNA molecules from two scrambled genes at different time intervals during macronuclear development in two ciliates, Stylonychia lemnae and Oxytricha trifallax. Our results revealed several shared characteristics of the gene unscrambling process in actin I orthologs in both species (Figure S6) and also in a different scrambled gene (TEBP) in Oxytricha trifallax. Firstly, our results showed that conventional IES deletion tends to occur before more complex rearrangement events. Secondly, during the gene unscrambling process, we found evidence to suggest that inversion, translocation or permutation events may occur through multiple parallel pathways, instead of following a deterministic order, although there may be preference for one or the other pathway in different species. Finally, among the surveyed molecules, we observed an unexpectedly high level of error at the boundaries flanking both conventional and scrambled IESs, sometimes associated with abnormal deletion, insertion, or permutation during development. Because most of this error is absent from the mature macronucleus, these observations are consistent with a mechanism that either eliminates or repairs the incorrectly rearranged strands. Together, our results suggest that some events take place in a preferred order, with removal of most nonscrambled IESs first, followed by the key events of DNA inversion and translocation/permutation.

Although an earlier study suggested that the process of IES elimination itself can be highly ordered in S. lemnae [39], we did not detect any specific linear order of conventional IES removal in S. lemnae actin I. The number of molecules analyzed, however, does not allow a precise test of correlation among IES excision order.

Conventional IES excision in Stylonychia and Oxytricha may be similar to the phenomena observed in Tetrahymena and Paramecium, where IESs are most likely marked by small RNAs, leading to modification of local chromatin structure [42,43] (reviewed in [31]). However, some IESs in Stylonychia and Oxytricha are smaller than the observed 26–27 nt small RNAs [29,32], which are also smaller than a nucleosome core particle. This poses a challenge, because small RNAs would be unable to mark such IESs precisely. This could be one source of error in IES excision. IES6 (scrambled between segments 8 and 10) in S. lemnae actin I may pose such a problem, as it is 17 bp flanked by a 6 bp pointer on one side and a 17 bp pointer on the other side, and this particular junction is involved in two examples of aberrant rearrangement at 6–8 bp cryptic pointers (cases f and i in Figure 5 and Table S2).

Because S. lemnae and O. trifallax eliminate the majority of their germline DNA (98% and 95%, respectively, with greater micronuclear genome size in S. lemnae), and then reorganize a substantial fraction of the DNA that remains, any mistakes of reorganization that yield nonfunctional genes could be fatal. Therefore, these processes must ultimately be very reliable and accurate, which is why the abundance of cryptic pointer junctions in our data was a surprise. Because most pointer sequences are too short to guide accurate IES removal and unscrambling of an entire gene, in part because of redundancy [15], other factors must assist unscrambling. The presence of either selective degradation or a repair mechanism is consistent with our observation that almost all molecules containing permutations have correctly eliminated all conventional IESs (with cases b and h in Figure 5 and cases a, b, and f in Figure S4A the only exceptions). The template-guided model offers a possible solution to these problems [33,34,36] with recent experimental support [36]. The proposed templates appear to be complete RNA copies of the DNA molecules present in the maternal macronucleus, appearing transiently in early macronuclear development [36]. The proposed use of maternal templates could simultaneously repair imprecise excision and guide unscrambling [36]. The presence of templates is also consistent with observations in Paramecium and Tetrahymena of an epigenetic influence of the old macronucleus on DNA rearrangements in the offspring [28,44–47]. This model is compatible with our conclusion that multiple unscrambling pathways may produce functional macronuclear molecules; however, it does not predict the distinct stages that we see during gene unscrambling.

We propose a general mechanism for gene unscrambling involving: marking of DNA sequences for deletion, recombination at adjacent or nearby repeats (that may correctly or incorrectly define segment boundaries), RNA template-guided DNA proof-reading of reparable errors or elimination of dead-end molecules...
containing grossly incorrect junctions, and finally, RNA template-guided recombination at scrambled pointers, yielding correct segment permutation and orientation.

With the exception of one IES in *O. trifallax actin I*, the examined molecules in this study eliminate all conventional IESs before they begin the more complicated steps of inversion and translocation. In *S. lemmnae actin I*, only after excision of all conventional IESs do we see evidence for inversion of segments 2 and 1 (or segments 3–10) and translocation of segments 1, 2 and 9. This strategy should increase the system’s robustness by reducing opportunity for unscrambling error. It is unknown whether rearrangement of other genes and their scrambled and nonscrambled orthologs in other species proceeds in a similar fashion. Each example may follow a precise pattern influenced by higher-ordered structures of genetic organization [18]. However, our analysis of both orthologous and unrelated genes in two different species consistently suggests that most conventional IESs are eliminated before DNA permutation. Therefore, this may reflect a general mechanism rather than a gene- or species-specific pattern.

The phenomenon of gene unscrambling in stichotrichous ciliates provides a unique model system to examine some of nature’s most ornate developmental DNA manipulations. More generally it provides a model system to study epigenetic influences on programmed genome rearrangement [36]. The *Oxytricha trifallax* genome project [48] will ultimately permit a detailed comparison of the *O. trifallax* germline and somatic genomes, revealing thousands of scrambled genes, many as complex as DNA polymerase α [15], and significantly more complex than the actin I locus.

**Materials and Methods**

**Cells and DNA**

*Stylonychia lemnnae* were grown in Pringsheim solution and fed daily with Chlorogonium elongatum [49]. To achieve conjugation, cells of different mating types (strain D9 and B1) were mixed and the stages of macronuclear development were determined by the size of macronuclear anlage. *Oxytricha trifallax* cells were treated as in [36]. Mating efficiency in both species was over 80% in all experiments and exconjugant survival rate was above 85%. *S. lemmnae* anlagen DNA was typically isolated from 5000–10,000 cells at different stages of macronuclear development (Figure 2) and used in PCR amplification as described in [37] and online Methods S1. Due to asynchronous differentiation of individual cells, each time point used in PCR comprises a window of approximately 4.5 hours, which was consistent across experiments in both species.

We collected DNA from conjugating *O. trifallax* cells at 0, 10, 25, 40, 48 and 55 hours after mixing cells of two mating types (strains JRB310 and JRB510). According to a microscopic survey using DAPI stain (not shown), *O. trifallax* conjugating cells reach the peak of the polytenation stage ~44 hours after mixing and enter the DNA poor stage ~55 hours after mixing, sharing a similar developmental time scale with *S. lemnnae* (polytenation peaks at 42 hours and the DNA poor stage starts ~52 hours; see Figure 2). Therefore, we label the time points in *O. trifallax* as follows: mixing, early, middle, late-a, late-b and DNA poor stages.

**PCR procedures**

DNA sequences flanking the micronuclear *S. lemnnae actin I* locus were recovered by UFW PCR [50] as in [25] using primers listed in Methods S1. All *S. lemnnae* partially-processed molecules were recovered by nested PCR or a second round of amplification. *O. trifallax* products were recovered in a single conventional PCR, except for one *TEBPα* product amplified with nested IES primers (Methods S1). Negative controls were always performed with water or macronuclear DNA as initial template. See [37] and Methods S1 for more details. Figure S7 provides sample agarose gel images of the PCR time-course for each gene and species.

**DNA cloning and sequencing**

*S. lemnnae* PCR products were purified (MinElute™ Gel Extraction Kit, Qiagen GmbH, Hilden, Germany) and cloned into pGEM®-T Easy vector (Promega GmbH, Mannheim, Germany). Plasmid DNA was isolated with the QIAprep® Spin Miniprep Kit (Qiagen GmbH, Hilden, Germany) and sequenced by MWG Biotech (Ebersberg, Germany) with T7 and SP6 primers. *O. trifallax* PCR products were purified (QiaQuick PCR Purification Kit, Qiagen, California, USA) and cloned into pCR2.1-topo vector (Invitrogen, California, USA). Plasmid DNA was isolated with the QIAprep® Spin Miniprep Kit (Qiagen, California, USA) and sequenced by Genezis (New Jersey, USA) or Agencourt (Massachusetts, Germany) with M13 forward and reverse primers.

**Sequence analysis**

BioEdit was used for alignment, and computational unscrambling of input sequences was performed with Gene Unscrambler [51].

**Supporting Information**

Figure S1 The positions of *S. lemnnae actin I* PCR primers and all primer pairs used to amplify molecules with and without permutations. Solid triangles are contained within MDS sequences and empty triangles derive from IES or flanking regions. Found at: doi:10.1371/journal.pone.0002330.s001 (0.16 MB DOC)

Figure S2 Details of the authentic pointers and cryptic pointers found in partially processed *S. lemnnae actin I* molecules with inaccurate deletions. A) DNA excision events around each IES. The first line of each set shows the precursor micronuclear sequences at each IES boundary. MDS sequences are in upper case, IES sequences in lower case and authentic pointers in bold upper case. The second line shows correct macronuclear excision products at authentic pointers (shaded red). The remaining set of lines shows the “sloppy” excision events at cryptic pointers (shaded pink). Numbers to the right of each line give the number of times a particular sequence was recovered. For IES 7 and 9, which are unconventional IESs that contain a telomere addition site on one side and a scrambled pointer on the other, the telomere locations are noted. B) Some PCR products contain large deletions that span several MDSs. These molecules can be grouped into five classes (o–5, also annotated in Figure 3) based on their deletion boundaries. Cryptic pointer-like sequences that flank these large deletions are shaded pink. These molecules are more likely to be dead-end products during erroneous DNA excision than repairable developmental intermediates. Annotation as in previous figures. Found at: doi:10.1371/journal.pone.0002330.s002 (1.33 MB DOC)

Figure S3 A schematic representation of all partially-processed *O. trifallax* (A) actin I and (B) *TEBPα* molecules involving conventional DNA deletion junctions at different stages of development (middle = 25 hr, late-a = 40 hr, late-b = 48 hr). MDS regions are blue, IESs yellow and authentic pointers maroon. Cryptic pointers used in some molecules are orange and highlighted by orange triangles. Primer pairs are shown under...
the schematic micronuclear map in each panel. Solid triangles represent MDS-specific primers; open triangles are IES-specific primers. Triangle direction indicates the strand polarity of the primers. Primer pairs in gray yielded only micronuclear-specific products at all developmental stages. Redundancy of the sequences is shown in parentheses. Aligned sequences are provided in Data S1 files “OtActinIConsventional.fas” and “OtTEBPaConventional.fas”, respectively. At the bottom of each panel is a summary of the locations of all cryptic pointers, indicated by orange flags, detected in all partially processed molecules (shown above and in Figure S4). Flags above the molecule represent cryptic pointers associated with aberrant deletions that do not alter micronuclear order. Flags below the molecule represent cryptic pointers found at aberrantly reordered junctions. Flag size is proportional to the number of times a cryptic pointer is observed at a specific location.

**Figure S4** Schematic representation of all partially-processed *O. trifallax* (A) *actin I* and (B) *TEBPa* molecules involving permutations. Molecules that do not contain any aberrant deletions or incorrect rearrangement are listed on the left as potential intermediates. Molecules with either aberrant deletions or incorrect permutations are provided on the right. MDS regions are blue, IESs yellow and authentic pointers in maroon. Cryptic pointers are orange and marked by orange triangles. Primer pairs are shown under each micronuclear map. Sold triangles represent MDS-specific primers; open triangles represent IES-specific primers. Triangle direction indicates the strand polarity of the primers. Redundancy of the sequences is shown in parentheses. Figure S3 provides a summary of all cryptic pointers used. All partially permuted sequences in this figure are provided in Data S1 files “OtActinIPermuted.fas” and “OtTEBPaPermuted.fas”.

Found at: doi:10.1371/journal.pone.0002330.s003 (0.11 MB DOC)

**Figure S5** A dot-plot-like representation of all isolated aberrantly rearranged *S. lemnae* *actin I* sequences (summarized in Figure 5). The micronuclear structure is represented vertically on the Y-axis. Each molecule is represented horizontally on the X-axis. The matched regions are showed as diagonal lines. The sequences are color-coded in the same way as in Figure 1. The gray lines on the X-axis represent sequences that may be included in the first round of PCR before nested PCR.

Found at: doi:10.1371/journal.pone.0002330.s005 (1.78 MB PDF)

**Figure S6** Schematic alignment of the orthologous *actin I* macronuclear and micronuclear sequences in *S. lemnae* and *O. trifallax*. Gray areas indicate alignable coding regions.

Found at: doi:10.1371/journal.pone.0002330.s006 (0.05 MB DOC)

**Figure S7** Representative Ethidium Bromide stained agarose gel images from analysis of PCR reactions that detected partially-processed products for (A) *S. lemnae* *actin I*, (B) *O. trifallax* *actin I* and (C) *O. trifallax* *TEBPa*, at different development stages. In (A), the predominant bands are micronuclear-specific products, and the processed products are indicated by open arrow-heads. M: marker; MAC: macronuclear DNA template control; NT: no template control. Primer combinations are shown beneath each micronuclear map.

Found at: doi:10.1371/journal.pone.0002330.s007 (0.30 MB DOC)

### Table S1

A summary of PCR results from different primer pairs at different developmental stages for (A) *S. lemnae* *actin I*, (B) *O. trifallax* *actin I* and (C) *O. trifallax* *TEBPa*. The ability of each primer pair to amplify micronuclear (MIC) and macronuclear (MAC) genomic sequences is indicated. The “+” and “−” signs indicate whether partially processed products can be observed among the PCR products, based on either agarose gel analysis or sequencing results.

Found at: doi:10.1371/journal.pone.0002330.s008 (0.09 MB DOC)

### Table S2

All cryptic pointers found in *S. lemnae* *actin I* molecules with unusual deletions or aberrant rearrangements. These are grouped based on whether the molecules have (A) deletion without permutation or (B) deletions with aberrant permutation (wrong MDS order). For the cryptic pointers in the first category, (A) lists the deleted IESs (all or partial), the MDS segments that are fused at the cryptic pointer, the number of times that each junction was observed, and the authentic 5’ and 3’ pointers that lie closest to the cryptic pointers in the germline sequence. For the cryptic pointers at aberrantly ordered junctions, (B) lists the observed rearrangements (joining MDS segments x to y), the number of times that each molecule type was observed, and the authentic 5’ and 3’ pointers that the cryptic pointers replace. In both (A) and (B), underlined nucleotides in the cryptic pointers are contained as a sub-string of the authentic (replaced) pointer, or vice-versa. Double underline is used for overlap when more than one cryptic pointer is a substring of the same actual pointer. Boldfaced nucleotides also overlap in position, i.e. the cryptic pointer used all or part of a real pointer on at least one side of the junction. In cases indicated by *, the actual position is a telomere addition site; therefore there are no neighbouring authentic pointers.

Found at: doi:10.1371/journal.pone.0002330.s009 (0.08 MB DOC)

### Table S3

Robustness analysis of *S. lemnae* *actin I* nonscrambled pointers (at conventional junctions). The lengths of the pointer and the IES (\(|\cdot|\)) between MDS x and y are listed. \(N_{\text{correct}}\) number of junctions that are covered in the assayed sequences; \(N_{\text{excised}}\) number of junctions with an excision event; \(N_{\text{correct}}\) number of junctions with an excision event at the correct pointer. *IES length excludes pointers. †Assuming that the accuracy of excision at different boundaries is independent, and that the error rates are not highly biased, we can roughly estimate the fraction of molecules that would be correctly-processed at all conventional IES sites by multiplying the values \((N_{\text{correct}}/N_{\text{excised}})\) in the last row of Table S3. Based on this approximation, most (~78%) molecules might be expected to contain at least one incorrect deletion event during development.

Found at: doi:10.1371/journal.pone.0002330.s010 (0.03 MB DOC)

### Table S4

Robustness analysis of *S. lemnae* *actin I* scrambled pointers at permuted junctions. Lengths of scrambled pointers between MDS x and y are listed. \(N_{\text{correct}}\) number of junctions with a rearrangement event; \(N_{\text{correct}}\) number of junctions with a rearrangement event at the correct pointer.

Found at: doi:10.1371/journal.pone.0002330.s011 (0.03 MB DOC)

### Method S1

Found at: doi:10.1371/journal.pone.0002330.s012 (0.05 MB DOC)

### Data S1

Found at: doi:10.1371/journal.pone.0002330.s013 (0.02 MB ZIP)
Acknowledgments

We thank Sabine Feiler for technical assistance and Glenn Herrick, Larry Klobbutcher, Mark Daley, Lila Kari, David Ardell, Dan Hogan, Michelle Dulbois, David Prescott, and all members of the Landweber and Lippis labs for discussions.

References

1. Gerbi SA (1986) Unusual chromosome movements in sciarid flies. Results Probl Cell Differ 13: 71–104.
2. Beermann S (1977) The diminution of Heterochromatogenous segments in Cyclops (Crustacea, Copepoda). Chromosoma 60(4): 297–344.
3. Tobler H (1969) The differentiation of germ and somatic cell lines in nematodes. Results Probl Cell Differ 13: 1–69.
4. Müller F, Tobler H (2000) Chromatin diminution in the parasitic nematodes ascarsis suum and parasarcis unicarial. Int J Parasitol 30(4): 391–399.
5. Harimann W, Volk H, Defranoux N, Wabir M (1993) Immunoglobulin class switch recombination. Annu Rev Immunol 11: 361–384.
6. Bassing CH, Swat W, Alt FW (2002) The mechanism and regulation of chromosomal V(D)J recombination. Cell 109 Suppl: S45–55.
7. Liher M (1996) Immunoglobulin diversity: rearranging by cutting and repairing. Curr Biol 6(2): 134–136.
8. Lewis SM (1994) The mechanism of VDJJ joining: lessons from molecular, immunological, and comparative analyses. Adv Immunol 66: 77–150.
9. Herrsoshi N (1989) A regulatory hierarchy for cell specialization in yeast. Nature 342(6251): 749–757.
10. Haber JE (1998) Mating-type gene switching in Saccharomyces cerevisiae. J Cell Biol 141(1): 1–6.
11. Bost P (1991) Molecular genetics of antigenic variation. Immunol Today 12(3): A29–33.
12. Vanhamme L, Pays E, Mc Cullogh R, Barry JD (2001) An update on antigenic variation in African trypanosomes. Trends Parasitol 17(7): 338–343.
13. Prescott DM (1994) The DNA of ciliated protozoa. Microbiol Rev 50(2): 253–267.
14. Mochizuki K, Gorovsky MA (2004) Small RNAs in genome rearrangement in Paramecium tetraurelia. Mol Cell Biol 24(17): 7370–7379.
15. Cavalcanti ARO, Landweber LF (2005) Insights into a biological computer: Detangling scrambled genes in ciliates. In: Chen J, Jonoska N, Rosenberg G, eds. 2005 Nanotechnology: Science and Computation. Berlin: Springer-Verlag, pp 349–360.
16. Cavalcanti AR, Clarke TH, Landweber LF (2005) MDS_IEs_DB: A database of macronuclear and micronuclear genes in spirotrichous ciliates. Nucleic Acids Res 33(Database issue): D396–398.
17. Cavalcanti AR, Clarke TH, Landweber LF (2005) DNA-mediated epigenetic programming of a genome rearrangement pathway. Nature 431(7075): 153–158.
18. Mollenbeck M, Cavalcanti AR, Jonsson F, Lippis HJ, Landweber LF (2006) Interconversion of germline-limited and somatic DNA in a scrambled gene. J Mol Evol 63(1): 69–73.
19. Chen J, Kadlifahr FF, Chen JZ (2007) DNA supercoiling suppresses real-time PCR: a new approach to the quantification of mitochondrial DNA damage and repair. Nucleic Acids Res 33(4): 1377–1380.
20. Wen J, Maaerter C, Lippis HJ (1996) Sequential excision of internal eliminated DNA sequences in the differentiating macronucleus of the hypotrich ciliate Stylonychia lemnacea. Nucleic Acids Res 24(22): 4413–4419.
21. Paschka AG, Hoeprich V, Jonsson F, Limdeke A, Weier G, et al. (2005) A microarray analysis of developmentally regulated genes during macronuclear differentiation in the stichotrichous ciliate Stylonychia lemnacea. Gene 359: 51–90.
22. Ehrenfeucht A, Petre I, Prescott DM, Rosenberg G (2002) String and graph reduction systems for gene assembly in ciliates. Mathematical Structures in Computer Science 12: 113–154.
23. Yao MC (1996) Programmed DNA deletions in Tetrahymena: mechanisms and implications. Trends Genet 12(1): 26–30.
24. Madireddy MT, Coyne RS, Smothers JF, Mickey KM, Yao MC, et al. (1996) Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell 87(1): 75–84.
25. Angleska A, Jonoska N, Saito M, Landweber LF (2007) RNA-guided DNA elimination in Tetrahymena. Cell 130(6): 841–851.
26. Paschka AG, Hoeprich V, Jonsson F, Limdeke A, Weier G, et al. (2005) A microarray analysis of developmentally regulated genes during macronuclear differentiation in the stichotrichous ciliate Stylonychia lemnacea. Gene 359: 51–90.  
28. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA (2002) Analysis of a pivi- related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 110(6): 689–699.
29. Juraneck SA, Rupprech S, Postherg J, Lippis HJ (2005) snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates. Eukaryot Cell 4(11): 1934–1941.

Author Contributions

Conceived and designed the experiments: LL, HL, YZ, AC MM BH. Performed the experiments: YZ MM FJ BH. Analyzed the data: LL, YZ AC MM FJ BH. Contributed reagents/materials/analysis tools: WC GR HL, YZ AC TD FJ SJ. Wrote the paper: LL, YZ MM BH.