Harms, Klaus; Lunnan, Asbjørn; Hülter, Nils; Mourier, Tobias; Vinner, Lasse; Andam, Cheryl P.; Marttinen, Pekka; Fridholm, Helena; Hansen, Anders Johannes; Hanage, William P.; Nielsen, Kaare Magne; Willerslev, Eske; Johnsen, Pal Jarle

**Substitutions of short heterologous DNA segments of intragenomic or extragenomic origins produce clustered genomic polymorphisms**

*Published in:* Proceedings of the National Academy of Sciences of the United States of America

*DOI:* 10.1073/pnas.1615819114

Published: 27/12/2016

**Document Version**
Publisher's PDF, also known as Version of record

**Published under the following license:**
CC BY-NC-ND

**Please cite the original version:**

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
Substitutions of short heterologous DNA segments of intragenomic or extragenomic origins produce clustered genomic polymorphisms

Klaus Harmsa,b,1, Asbjørn Lunnaan, Nils Hülterb, Tobias Mourierb, Lasse Vinnenb, Cheryl P. Andamg, Pekka Mattiinenb, Helena Fridholmb,f, Anders Johannes Hansenb,William P. Hanaged, Kaare Magne Nielsenh, Eske Willerslevb,1, and Pål Jarle Johnsena,1

aDepartment of Pharmacy, Faculty of Health Sciences, The Arctic University of Norway, 9037 Tromsø, Norway; bCentre for GeoGenetics, Natural History Museum of Denmark; University of Copenhagen, 1350 Copenhagen K, Denmark; cGenomic Microbiology, Institute of Microbiology, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany; dDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115; eHelinski Institute for Information Technology, Department of Computer Science, Aalto University, FIN-00076 Aalto, Finland; fDepartment of Microbiological Diagnostics and Virology, Statens Serum Institut, 2300 Copenhagen S, Denmark; gDepartment of Life Sciences and Health, Oslo and Akershus University College of Applied Sciences, 0130 Oslo, Norway; and hGenØk-Center for Biosafety, 9294 Tromsø, Norway

In a screen for unexplained mutation events we identified a previously unrecognized mechanism generating clustered DNA polymorphisms such as microindels and cumulative SNPs. The mechanism, short-patch double illegitimate recombination (SPDIR), facilitates short-single-stranded DNA molecules to invade and replace genomic DNA through two joint illegitimate recombination events. SPDIR is controlled by key components of the cellular genome maintenance machinery in the gram-negative bacterium Acinetobacter baumyi. The source DNA is primarily intragenomic but can also be acquired through horizontal gene transfer. The DNA replacements are non-reciprocal and locus independent. Bioinformatic approaches reveal occurrence of SPDIR events in the gram-positive human pathogen Streptococcus pneumoniae and in the human genome.

illegitimate recombination | mutation | microindels

Short patches of clustered nucleotide variations are routinely observed in whole genome comparisons (1, 2). These sequence variations are substrates for natural selection, which shapes prokaryotic (3, 4) and eukaryotic (5, 6) genomes. Clustered nucleotide variations also play a role in oncogenesis where they add to the overall genomic instability (7, 8). Despite their significant biological role, the molecular mechanisms underlying formation of clustered nucleotide variations are not fully understood.

Known mechanisms responsible for clustered nucleotide variations include error-prone DNA polymerases (9) and conversions at imperfect palindromes through template-switching (10) (templated mutagenesis), which can generate tracts of single nucleotide changes, respectively. Down-regulation or loss of genes involved in mismatch repair can also lead to increased genome-wide point mutation frequencies that can result in random single-nucleotide variation (SNV) clusters. Moreover, cumulative SNVs have been described when genes for DNA-modifying enzymes were up-regulated (11). All these mechanisms typically result in tracts of single-nucleotide polymorphisms (SNPs).

More complex clustered genomic polymorphisms may also develop through point mutations accumulating in a small DNA tract over a short time or through independent insertion and deletion events (12). A number of RecA-independent mechanisms have been described and investigated in detail that lead to microdeletions without insertions, or to microinscriptions without deletions, in both prokaryotic and eukaryotic organisms. Among these mechanisms are replication slippage (13) or copy number variations in microsatellite DNA (14), illegitimate recombination at microhomologies (15, 16), imprecise nonhomologous end joining (NHEJ) (17), DNA gyrase-mediated strand switching (18), and transposon scars. Two or more temporally independent deletion/insertion events at the same locus can result in clustered polymorphisms, although in retrospective studies, such sequential events are nearly impossible to verify.

The most diverse clusters of nucleotide variations are formed by microhomology-mediated end-joining (MMEJ). MMEJ has been observed in eukaryotes only and can repair DNA double-strand (ds) breaks in an error-prone way. During repair, MMEJ often generates short, direct, or inverted repeats (19) and occasionally incorporates ectopic DNA at the recombinant joints (20). MMEJ results in highly variable clustered polymorphisms at the recombinant joint and is now recognized as a driving force in rapidly evolving oncogenic cells (21). DNA polymerase theta (POLQ) has recently been identified as the key enzyme in MMEJ-directed error-prone repair, but many mechanistic details of its function remain elusive (22). To date, no POLQ-like genes have been identified in prokaryotes.

Due to the immense evolutionary and biomedical implications of how and why genetic diversity is generated in prokaryotic and eukaryotic organisms, the underlying mechanisms are intensively investigated. To study and quantify the formation of clustered polymorphisms, we developed a detection assay in the bacterium

Significance

Clustered genomic polymorphisms in DNA, such as microindels and stretches of nucleotide changes, play an important role in genome evolution. Here, we report a mutation mechanism responsible for such genomic polymorphisms where short, single-stranded DNA molecules invade double-stranded DNA and replace short genomic segments. We show, in a bacterial model organism, that the genomic replacements occur with very low levels of sequence identity (microhomologies). The invading DNA can be of intergenic or foreign origin. Genotoxic stress, horizontally taken-up DNA, or lack of genome maintenance functions increase the mutation frequency up to 7,000-fold. Bioinformatic approaches suggest that this class of mutations is widespread in prokaryotes and eukaryotes and may have a role in tumorigenesis.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

FREely available online through the PNAS open access option.

1To whom correspondence may be addressed. Email: klaus.harms@spdir.net, ewillerslev@snm.ku.dk, or paal.johnsen@uit.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615819114/-/DCSupplemental.

This article is a PNAS Direct Submission.

FREely available online through the PNAS open access option.

1To whom correspondence may be addressed. Email: klaus.harms@spdir.net, ewillerslev@snm.ku.dk, or paal.johnsen@uit.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615819114/-/DCSupplemental.
Acinetobacter baylyi. We demonstrate how regions of clustered, highly variable DNA sequence variations (ranging from 3 to 77 bp) can be formed by two coupled, microhomology-dependent illegitimate recombination (IR) events with free DNA single strands of intragenomic or external origin.

Results

Joined Double Illegitimate Recombinations Generate Clustered Polymorphisms. To quantify and characterize clustered small indels and polymorphisms, we developed an in vivo detection construct (hisC::ND5′) (23) in the soil bacterium Acinetobacter baylyi ADP1. The construct is permissive for small IR events but largely refractory to single-nucleotide mutations. In this construct, two neighboring stop codons in a functionless 228-bp insert prevent expression of a histidine prototrophy marker gene (histidinol-phosphate aminotransferase; Fig. 1 D). We found that spontaneous histidine-prototrophic (His+) mutants arose at low frequencies. Subsequent DNA sequencing analyses of individual His+ isolates revealed that the ‘ND5′ segment was frequently substituted with different heterologous segments of intragenomic origins. The substituting DNA segments were of similar or shorter length, eliminating or bypassing the stop codons (Fig. 1 B–E and Dataset S1), and their neighboring upstream and downstream nucleotide stretches were identical with DNA segments in otherwise fully heterologous DNA regions elsewhere in the genome (Fig. S1). Sequence analyses of these donor DNA segments and the parental DNA sequences strongly suggested that integration occurred through hybridization at microhomologies (short identical DNA stretches) or at extended microhomologies (clusters of microhomologies interrupted by mismatches and gaps in heterologous DNA; Fig. 1 B–E and Supporting Information) followed by illegitimate recombinations.

The recombinations occurred either at a single, contiguous microhomology (class 1 events; Fig. 1 B and C) or at two separate microhomologies on the same molecule (class 2 events; Fig. 1 D and E). The recombinations were nonreciprocal (Supporting Information) and independent of genomic locus and detection construct (Supporting Information). Together, these short-patch double illegitimate recombination (SPDIR) events led to highly variable polymorphisms at a single genetic locus and introduced multiple clustered nucleotide exchanges, DNA sequence replacements of variable length, or deletions accompanied by nucleotide changes at the deletion site (Dataset S1), resulting in highly diverse codon changes (Fig. S2). In all characterized SPDIR events, the source DNA of the acquired nucleotide polymorphisms was identified both for intragenomic and extragenomic (see below) origins (Dataset S1). Net nucleotide gains (maximum six base pairs) were observed in only a few cases. Although the SPDIR mechanism depends on microhomologies, the randomness of the genetic changes observed suggests a broad mutagenic potential.

Low Frequency of SPDIR Mutations in Wild-Type Cells. We quantified occurrence of SPDIR experimentally in wild-type (WT) A. baylyi cells and found that His+ revertants were scarce (1.1 × 10−11, about 14-fold rarer than single point mutations; Table 1). The fraction of SPDIR mutation events among the His+ reversions was ~2%, corresponding to a calculated SPDIR frequency of 5.6 × 10−13 (Table 1). This number is likely an underestimation due to limitations in the detection construct because SPDIR-generated substitutions that introduce stop codons or frameshifts or lead to improper protein folding may remain undetected.

The non-SPDIR His+ mutations were in most cases (>90%) in WT conferred by frame deletions in ‘ND5′ [i.e., single illegitimate recombination (IR) events], both with and without microhomologies, and occasionally by different classes of mutations (Supporting Information). The fact that SPDIR occurred in the WT close to the detection limit in our specific experimental setup can explain lack of prior experimental discovery.

Single-Strand–Specific DNA Exonucleases Control SPDIR in Wild-Type Cells. Microhomology-mediated IR events have been observed in prokaryotes and eukaryotes (15, 16) and are initiated by annealing of DNA single-strand ends. We hypothesized that SPDIR was initiated by hybridization of genomic dsDNA at exposed single-stranded (ss) gaps, loops, or replication forks, with ssDNA segments. In prokaryotes, free cytoplasmic DNA single strands are attacked by ss-specific DNA exonucleases (24) (ssExo), and in A. baylyi, these ssExo have been revealed as RecJ and ExoX (23). We therefore quantified SPDIR in ssExo-deficient mutants and found that the SPDIR frequency was elevated approximately sevenfold in ΔrecJ and fourfold in ΔexoX mutants (Table 1). The frequency was increased 28-fold in a ΔrecJ ΔexoX double mutant, which lacked all ssExo activity. In the ΔrecJ ΔexoX strain, SPDIR events produced about 34% of all His+ mutation events, whereas in WT and in the single mutants the proportion of SPDIR events was at least sixfold lower than in the ΔrecJ ΔexoX mutant (Table 1). These results confirmed that SPDIR is suppressed by ssExo in WT cells and indicate that SPDIR events depend on the presence of ssDNA in the cytoplasm.

SPDIR Is Inhibited by RecA Protein. Cytoplasmic ssDNA is a cellular genome damage signal and can be bound by RecA protein to initiate recombinational repair and to trigger the SOS response (25). We deleted the recA gene of A. baylyi, and in the ΔrecA mutant we observed an about sixfold SPDIR frequency increase. Remarkably, in a ΔrecA ΔΔrecJ ΔexoX triple mutant, the SPDIR frequency was >7,700-fold higher than that of the WT, and SPDIR was the most common His+ mutation (80%; Table 1). The strong synergy effect suggests that SPDIR is controlled by factors beyond elimination of free cytoplasmic DNA. It is conceivable that binding of RecA protein to ssDNA efficiently prevents hybridization of ssDNA molecules, and molecules that escape RecA-binding frequently anneal at microhomologies. In WT cells, these microhomology- annealed molecules are attacked.

![Fig. 1](https://example.com/fig1.png)
by ssExo and prevented from genomic integration, as observed in *Escherichia coli* (24) and *A. baylyi* (23). Alternatively, faithful recombinational DNA damage repair mediated by RecA together with ssExo prevents production of ssDNA remnants (26) (e.g., displaced strand fragments or flaps) that could act as donor molecules for SPDIR. These explanations are not mutually exclusive.

**Exposure to Genotoxic Stress Increases SPDIR Frequencies.** IR frequencies are increased with accumulating genomic DNA damages, and the increase has been attributed to microhomology-mediated DNA end-joining events leading to deletions and other genomic rearrangements (27). We determined whether introduction of foreign DNA isolated from *Bacillus subtilis* resulted in a fourfold to fivefold elevated SPDIR frequency (Table 1). We repeated the experiments with the ∆*recJ* ∆*exoX* mutant, using *B. subtilis* isogenic *A. baylyi* His" DNA, and DNA isolated from salmon sperm as donor DNA substrates. In the ∆*recJ* ∆*exoX* strain, addition of the DNA substrates led to SPDIR frequencies about 15- to 40-fold higher than without added DNA (Table 1). Notably, when exposed to foreign DNA, about two thirds of the SPDIR mutations were formed with cognate DNA, and approximately one third were formed with taken-up DNA. This result is consistent with findings of previous reports showing that recombination attempts during natural transformation frequently result in DNA strand breaks and thus can damage genomic DNA (32). The DNA damages then lead to increased SPDIR frequencies, as observed in the experiments with ciprofloxacin and UV light. The RecA-independent recombination at the MH was strand orientation-specific (Supporting Information). In a transformation-deficient ∆*comA* ∆*recJ* ∆*exoX* triple mutant [lacking the ComA DNA uptake pore (23)], the SPDIR frequency was not different from that of the ∆*recJ* ∆*exoX* mutant (Table 1). These results confirm that SPDIR is primarily an intragenomic process and also demonstrate that natural transformation can be mutagenic through the SPDIR pathway. Consequently, clustered polymorphisms in the genome of some bacterial species can be the result of foreign DNA acquisition. However, in retrospective genome analyses it may often not be possible to identify the origin of donor DNA molecules due to the short length of the SPDIR-generated polymorphisms.

By contrast, the result of foreign DNA acquisition. However, in retrospective genome analyses it may often not be possible to identify the origin of donor DNA molecules due to the short length of the SPDIR-generated polymorphisms.

**Natural Transformation Increases Frequency and Variability of SPDIR Events.** To explore the effect of exogenous DNA on SPDIR formation, we exploited the constitutive competence for natural transformation of WT *A. baylyi* cells (23). DNA molecules are taken up by the cells into the cytoplasm as single strands (31). We found that exposure to foreign DNA isolated from *Bacillus subtilis* resulted in a fourfold to fivefold elevated SPDIR frequency (Table 1). We repeated the experiments with the ∆*recJ* ∆*exoX* mutant, using *B. subtilis* isogenic *A. baylyi* His" DNA, and DNA isolated from salmon sperm as donor DNA substrates. In the ∆*recJ* ∆*exoX* strain, addition of the DNA substrates led to SPDIR frequencies about 15- to 40-fold higher than without added DNA (Table 1). Notably, when exposed to foreign DNA, about two thirds of the SPDIR mutations were formed with cognate DNA, and approximately one third were formed with taken-up DNA. This result is consistent with findings of previous reports showing that recombination attempts during natural transformation frequently result in DNA strand breaks and thus can damage genomic DNA (32). The DNA damages then lead to increased SPDIR frequencies, as observed in the experiments with ciprofloxacin and UV light. The RecA-independent recombination at the MH was strand orientation-specific (Supporting Information). In a transformation-deficient ∆*comA* ∆*recJ* ∆*exoX* triple mutant [lacking the ComA DNA uptake pore (23)], the SPDIR frequency was not different from that of the ∆*recJ* ∆*exoX* mutant (Table 1). These results confirm that SPDIR is primarily an intragenomic process and also demonstrate that natural transformation can be mutagenic through the SPDIR pathway. Consequently, clustered polymorphisms in the genome of some bacterial species can be the result of foreign DNA acquisition. However, in retrospective genome analyses it may often not be possible to identify the origin of donor DNA molecules due to the short length of the SPDIR-generated polymorphisms.

**The Two IR Events of SPDIR Are Temporally Linked.** Three lines of evidence strongly suggest that SPDIR mutations form within a single generation before selection. First, we frequently found

---

### Table 1. His" and SPDIR frequencies in *A. baylyi* strains without and with genotoxic stress or addition of DNA

<table>
<thead>
<tr>
<th><em>A. baylyi</em> ADP1 hisC::ND5′</th>
<th>SPDIR fraction</th>
<th>Calculated SPDIR frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIP₄ (MIC)</strong></td>
<td><strong>UV, mJ₂₆₀ nm</strong></td>
<td><strong>DNA</strong></td>
</tr>
<tr>
<td>Wild type</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wild type</td>
<td>0.1</td>
<td>—</td>
</tr>
<tr>
<td>0.25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wild type</td>
<td>—</td>
<td>10.8</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>3.6</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>10.8</td>
</tr>
<tr>
<td>Wild type</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>∆<em>recJ</em> ∆<em>exoX</em> ∆<em>comA</em></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>his&quot; <em>trpE</em>72</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*P. n. a., not applicable.*

*Supplemented with 300 ng mL⁻¹ genomic DNA from the following sources: BS, *Bacillus subtilis* 168; AB, *A. baylyi* hisC::ND5′; SS, salmon sperm DNA.

Identical genotypes were regarded as siblings originating from a single mutation event.

The SPDIR events formed with endogeneous AB DNA.

Eight SPDIR events were formed with BS, and 15 events were formed with AB DNA. One donor DNA segment was present in both donor genomes.

Point mutation frequency, given as median Trp⁷ frequency.
intragenomic donor DNA segments in SPDIR isolates with reverse complement orientation relative to the hisC::’ND5i’ allele (Table S1). In these cases, temporally independent IR events would result in lariat chromosome intermediates that cannot be replicated by the cell. Second, SPDIR events with foreign DNA (that is taken up by the cell as ssDNA fragments) require genomic integration through two IR events in a single generation to prevent potentially lethal dsDNA breaks. Third, many His' recombinants from the same primary cultures frequently carried unique, identical SPDIR mutations, both with intragenomic or exogenous donor DNA molecules. Such jackpot events strongly suggest that SPDIR mutations preexisted in the bulk culture (see Supporting Information and Table S1 for details). In our frequency calculations, identical mutations from the same assay were treated as single mutation events.

A Model for SPDIR Caused by Cytoplasmic ssDNA Molecules. We show that SPDIR events depend on the presence of ssDNA and are suppressed by key components of the genome maintenance machinery. A genomic integration model is depicted in Fig. 2. In that model, microhomologies are used by the cell to join unrelated DNA molecule ends, as has been demonstrated and quantified in previous studies for single (15) or multiple (24, 33) IR events. The model further builds on a proposed mechanism for strand orientation-specific, RecA-independent integration of short DNA molecules (23), in which we showed that fully homologous oligodeoxynucleotides (≥200 bp) could be chromosomally integrated in a single event during replication, acting as primers for Okazaki fragments (23, 24) (Supporting Information). In the present study, we demonstrate that microhomologies are sufficient for chromosomal integration at low but detectable frequencies during lagging strand DNA synthesis (Fig. 2, Fig. S3, and Supporting Information).

Bioinformatic Analyses Reveal Putative SPDIR Events in Human Genomes. For humans, we isolated DNA from blood samples and colon cancer tissues from three individuals (36) and sequenced the DNA on an Illumina HiSeq 2000. We called clustered polymorphisms with donor molecules for SPDIR largely as described above (see Supporting Information for details). Altogether, we identified 94 putative SPDIR events (Table S2 and Dataset S3). Detailed analyses showed that more than half of these events were short clustered nucleotide variations present in various human sequence databases including alternative genome assemblies, suggesting that SPDIR contributes to the generation of human heterozygous alleles and that SPDIR is a mutation mechanism operative in humans. The remaining insertion–deletion sequences were not found in available databases and were considered novel (Table S2). Seven novel putative SPDIR events were uniquely identified in DNA from blood, whereas a total of 33 putative novel events were identified in DNA from colon cancer tissue only (Table 2). Remarkably, 16 novel SPDIR events from cancer tissue formed at predicted hairpins and led to microinversions (Fig. 3) that in two cases were imprecise (Fig. 3B). These microinversions predictively formed through donor ssDNA molecules that originated from the same locus but reannealed with the DNA single strand as reverse complement. Donor DNA molecules from loci very close to the SPDIR site were also observed in the experimental studies (4. baylyi SPDIR isolates A4 and K49; Dataset S1), but the hisC::’ND5i’ detection construct did not contain stem–loop secondary structures. Close proximity between donor locus and recombinant microindel locus may increase the likelihood for a SPDIR event.

Three novel SPDIR events, including a single microinversion, were identified both from cancer tissue and from blood (Table 2), suggesting somatic mutations early in embryogenesis, spread of genetic material within the body, or previously unknown heterozygous alleles. The observed predominance of SPDIR in colon cancer tissue possibly reflects the reduced activity of genome maintenance functions generally observed in cancer cells (37, 38). This observation is consistent with the increased SPDIR frequencies of genome maintenance mutants in our experimental bacterial system (Table 1).

Harms et al.
Table 2. Combined numbers of novel SPDIR events from three human individuals

<table>
<thead>
<tr>
<th>Putative novel SPDIR events</th>
<th>Cancer</th>
<th>Cancer and blood</th>
<th>Blood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>33</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Associated with genes</td>
<td>20</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>ORFs</td>
<td>15</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Potential control regions</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tumorigenesis</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Growth and proliferation, differentiation, apoptosis, DNA binding, and transcription</td>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Other functions</td>
<td>10</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Not associated with genes</td>
<td>13</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Microinversions</td>
<td>16</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The potential SPDIR numbers for each human individual are listed in Table S2.

Discussion

In this study we identified a previously unrecognized mechanism, SPDIR, which generates clustered DNA polymorphisms. We show that SPDIR facilitates the formation of SNP clusters, microindels, and mosaic genes (experimentally observed substitutional insertion of up to 26 codons; Dataset S1). SPDIR occurs by ssDNA segments of intragenomic or extragenomic origins that invade and replace genomic DNA through two IR events.

Our genetic studies in A. baylyi with specific deletion mutants, together with the genotoxic stress and transformation experiments, clearly show that cytoplasmic ssDNA segments are responsible for SPDIR (Fig. 2). In wild-type cells, cytoplasmic ssDNA is a genomic damage signal, and the formation of ssDNA is tightly controlled (25). SPDIR can be classified both as a recombination and as a replication-associated mutation mechanism for clustered polymorphisms, with rare ssDNA segments acting as mutagens. Although oligonucleotides are known to recombine intracellularly or in the course of horizontal gene transfer (23, 24), and synthetic oligonucleotides are now widely used in targeted mutagenesis approaches, these events are based on DNA homology. SPDIR depends exclusively on microhomologies in otherwise heterologous DNA that can be as short as 12 bp and interrupted by mismatches and gaps.

SPDIR occurs rarely in A. baylyi wild-type cells. However, DNA damages increase the SPDIR frequency by orders of magnitude. Consequently, the cells turn into transient phenotypic mutants for microindels under genotoxic stress. The transient mutator phenotype does not require mutations in DNA repair genes, as frequently observed in mismatch repair-deficient mutants of prokaryotes and eukaryotes (39). It is conceivable that increased SPDIR frequencies can provide cells with a competitive advantage in fluctuating environments, as reported for genotypic mutants (40, 41). SPDIR can generate near-random genetic variations and alter entire protein domains in a single generation. It is thus tempting to speculate that SPDIR may be an important mechanism in protein evolution (42) following gene amplification and duplication events (43) (Supporting Information).

Our in silico identification of potential SPDIR events in both the gram-positive pathogen S. pneumoniae and in the human genome strongly suggests that SPDIR is a general mutation mechanism with relevance beyond our model organism A. baylyi. The identified microindel variants, together with the presence of intragenomic donor molecules, are consistent with the experimentally obtained SPDIR events and thus biologically plausible. Typical SPDIR-generated sequence changes are inaccessible by known point mutation or recombination processes, such as replication slippage, microhomology-dependent IR, NHEJ, DNA gyrase-mediated strand switching, or transpositions. However, sequence variations caused by SPDIR are comparable with those produced by MMEJ, a highly mutagenic DNA repair mechanism in eukaryotes (20). MMEJ is tightly down-regulated in healthy cells but often operative in tumor tissue. DNA double-strand breaks are repaired by MMEJ in an error-prone way, frequently leading to incorporation of ectopic DNA segments at the joints (20).

In our human tumor samples, we determined that 16 uniquely identified clustered polymorphisms were microinversions at predicted hairpins (Fig. 3 and Dataset S3). Microinversions at hairpins have been reported (44–46), but the mechanistic details of their formation remain elusive (45) and are considered unrelated to templated mutagenesis at imperfect hairpins (46). The formation of microinversions is also not consistent with our current understanding of the MMEJ or of other mutation mechanisms, and microinversions at hairpins have not been reported in MMEJ surveys (19, 20). However, microinversions can be explained most parsimoniously by SPDIR where the inverted repeats of the hairpins act as microhomologies and are used for the illegitimate joints (Fig. 3A), consistent with the model shown in Fig. 2. Our results indicate that SPDIR-caused mutations occur in colon cancer at elevated frequencies but not in the whole blood control. In many cancers, including those with up-regulated MMEJ, genome maintenance mechanisms such as Rad51 (eukaryotic RecA homolog) and ssExo are down-regulated (47, 48). It is conceivable that SPDIR occurs at elevated frequencies in such tumor cells, as experimentally observed in the A. baylyi ΔrecA ΔrecF ΔexoXA triple mutant (Table 1). The role of SPDIR in cancer progression requires further exploration.

Materials and Methods

The A. baylyi mutant strains were constructed as described (23, 32, 49) with standard procedures (Supporting Information) and are listed in Table S3. The mutation experiments were conducted in liquid cultures that were inoculated with a single colony of a His strain and aerated for 15 h at 30 °C in LB broth. The cells were washed, plated on M9 minimal medium with 10 mM succinate (M9S; His" mutant titer) and in appropriate dilution on LB (total cell titer), and incubated at 30 °C. When applicable, ciprofloxacin or DNA was added before inoculation. When UV was used as DNA-damaging agent, the cells were grown for 11 h, washed in PBS, irradiated with a germicidal lamp, and then grown in LB for another 4 h. On the M9S selective plates, His" cells grow less than one generation.

His" colonies on M9S were picked after 40 h (recA strains) or 64 h (recAΔrecA strains) and restreaked on M9S, and the recombinant hisC segment was amplified by PCR and Sanger-sequenced (Supporting Information). To identify ectopic inserts, the sequencing results were aligned with the A. baylyi genome and, when donor DNA for natural transformation was used, with donor DNA sequences, using BLAST (50).

The bioinformatic approaches are described in detail in Supporting Information. The R scripts are available from the authors upon request.

Two ethical boards reviewed the protocol for investigation of the human samples included in this study: the Regional Committee on Health Research Ethics (Case H-2–2012-FSP2) and the National Committee on Health...


49. Seguin-Orlando A, et al. (2013) Ligation bias in Illumina next-generation DNA library preparation and the consequences for technical assistance. The cancer work was supported by The Danish National Research Foundation (K.H.), the Academy of Finland Grant 251170 (to P.M.), the Finnish Centre of Excellence in Computational Inference Research Grant 259272 (to P.M.), and also by Norwegian Research Council Grant 204263/F20 (to P.J.J.).