Open Journal of Genetics, 2014, 4, 227-238
Published Online June 2014 in SciRes. http://www.scirp.org/journal/ojgen
http://dx.doi.org/10.4236/ojgen.2014.43023

Scientific Research

Impact of Sequence Non-Identities on Recombination within the *pil* System of *Neisseria gonorrhoeae*

Stuart A. Hill*, Jenny Wachter

Department of Biological Sciences, Northern Illinois University, DeKalb, USA
Email: *sahill@niu.edu

Received 5 April 2014; revised 4 May 2014; accepted 3 June 2014

Abstract

Neisseria gonorrhoeae engages in extensive _intra-cellular_ gene conversion between the PilE-expression locus (_pilE_) and the transcriptionally-silent _pil_ gene copies (_pilS_). _In silico_ analyses were applied to investigate the extent of sequence heterogeneity between the various _pilS_ gene copies. Analysis of synonymous and non-synonymous substitutions between the different _pilS_ genes indicated that relatively few amino acid changes would occur due to nucleotide polymorphisms towards the 5' end of the _pilS_ genes whereas more frequent amino acid substitutions would be incorporated within the "hypervariable" region. The lack of non-synonymous substitutions at the 5' end of the genes was found to be under selective pressure as indicated by a positive D_T score utilizing the Tajima test. The presence or absence of mismatch repair appeared to only impact recombination when non-identical DNAs recombined via the DNA transformation route, where small _pil_ sequence heterogeneities were sufficient to terminate recombination tracts, with these sequence constraints being relieved in cells carrying a _mutS_ mutation. Therefore, the data indicate that the effect of sequence heterogeneity on recombination within the _pil_ system appears to depend upon the context with which the non-identical DNAs recombine.

Keywords

Recombination, Antigenic Variation, Pilin, Sequence Heterogeneity

1. Introduction

Neisseria gonorrhoeae (the gonococcus) causes the sexually transmitted disease gonorrhea, which is generally a

*Corresponding author.

How to cite this paper: Hill, S.A. and Wachter, J. (2014) Impact of Sequence Non-Identities on Recombination within the *pil* System of *Neisseria gonorrhoeae*. *Open Journal of Genetics*, **4**, 227-238. http://dx.doi.org/10.4236/ojgen.2014.43023

non-complicated mucosal infection of humans that is characterized by a massive neutrophil infiltration. Though treatable by antibiotic therapies, vaccine development has been hampered due to the inability of the host to mount an effective immune response, in part, due to extreme antigenic and phase variation of several different gonococcal surface components (e.g., PilE polypeptide, *opa* gene expression, LOS variation; reviewed [1]).

The major protein subunit of the pilus organelle, PilE polypeptide, is encoded by the *pilE* gene, and is capable undergoing antigenic variation *in vivo* [2] [3]. Besides *pilE*, the chromosome also contains other variable *pil* gene sequences in various silent loci (*pilS*) [4] [5]. *pilS* loci contain multiple *pil* gene copies arranged in a tandem array. The variable *pilS* gene copies differ from the *pilE* gene in that they are believed to be non-transcribed due to a lack of promoter elements, as well as lacking the conserved 5' 150 bp found in all *pilE* genes. Despite these differences, all *pil* genes have common characteristics with variable segments being interspersed with small constant regions [6]. As the gonococcal chromosome contains multiple *pil* gene sequences (the number varies depending upon the strain), considerable sequence heterogeneity exists within the variable gene segments. Consequently, *pilE/pilS* recombination is a classic example of homeologous recombination.

PilE antigenic variants are produced following *pilE* recombination with a *pilS* gene copy (schematic model is presented in **Figure 1**). This *intra*-cellular event occurs as a RecA-dependent, non-reciprocal gene conversion with information being transferred uni-directionally from *pilS* to *pilE* with loss of the corresponding *pilE* gene sequence during the gene conversion event [4] [7]-[10]. Recently, a unique DNA structure has been identified at the *pilE* locus that has been proposed to initiate *pilE/pilS* recombination [11]. The formation of this guanine quartet structure apparently allows a nick to be introduced which is then acted upon by RecJ exonuclease to yield single-stranded DNAs that can then recombine with a *pilS* locus to yield *pilE* gene variants.

The release of the *N. gonorrhoeae* FA1090 genome sequence (GenBank accession number AE004969), followed by the release of numerous other *Neisseria* genome sequences, identified many of the recombination and repair proteins that have been described for *Escherichia coli* [12]. In particular, the RecA, RecB, RecC, and RecD proteins as well as a partial complement of the RecF pathway components. Gonococci also contain partial complements of other repair systems described for *E. coli*. They appear to utilize mismatch repair as they possess MutS and MutL homologues, yet lack the *mutH* and the *dam* methylase genes [13] [14]. Consequently, mismatch repair of single base mismatches or short insertions or deletions in the *Neisseriae* does not appear to be methyl-directed.

In this study, we examine, *in silico*, the extent of *pil* sequence heterogeneity within the *pilS* gene copies and determine its effect on the *pil* recombination system. The analysis demonstrates considerable sequence heterogeneity between the *pilS* gene copies, yet apparent selective pressure constrains 5' nucleotide polymorphisms within *pilS* that would allow only synonymous amino acid changes to occur within this region of *pilE* following a *pilE/pilS* recombination events. In contrast, nucleotide substitutions leading to non-synonymous amino acid changes are predicted to occur towards the 3' end of the gene within the so-called "hyper-variable" region. Previously, we had shown that the presence or absence of an active mismatch repair system did not seemingly affect the efficiency of *intra*-cellular *pilE/pilS* recombination [15]. However, it appears that short sequence non-identities between variant *pil* segments impede DNA transformation-mediated recombination in *rec*+ bacteria, with such constraints being alleviated through inactivation of the mismatch repair system. Thus, the data indicate that the manner by which *pil* heteroduplexes are created dictates how the cell deals with recombination between non-identical *pil* DNAs.

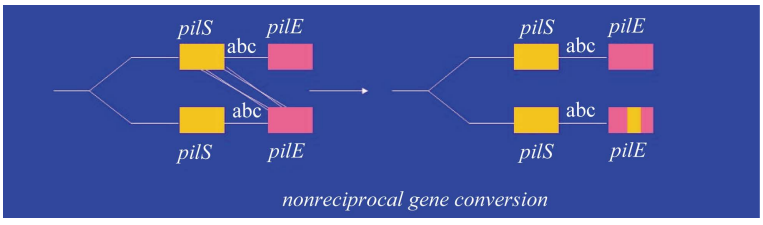

Figure 1. Schematic model of *pilE/pilS* recombination in *N. gonorrhoeae* *pilE* recombines with a *pilS* locus through a mechanism that resembles non-reciprocal gene conversion. Genetic information is uni-directionally tranferred from the *pilS* locus to *pilE* resulting in a variant *pilE* gene and an unchanged *pilS* gene copy.

2. Materials and Methods

2.1. Strains and Growth Conditions

Neisseria gonorrhoeae strain MS11 was used for all aspects of this study. Gonococci were grown on gonococcal typing medium [16] at 37˚C in a 5% CO_2 atmosphere. When grown in liquid culture 420 ng/ml $NaHCO_3$ was added to the culture medium and the agar was omitted. Where appropriate, antibiotics were added at the following concentrations; chloramphenicol 10 μg/ml, kanamycin 80 μg/ml, and erythromycin 5 μg/ml. The various *pilE* variants and mutants have previously been described [15] [17] [18]. Colony morphology variants were assessed using a phase contrast microscope as previously described [2] [8]. The *recB* growth suppressor mutant arose on plates and has previously been described [19].

All cloning was performed with *Escherichia coli* DH 5α using standard protocols. Transformants were selected on plates containing antibiotics at the following concentrations; ampicillin, 100 μg/ml, erythromycin, 200 μg/ml, kanamycin 80 μg/ml, and chloramphenicol, 20 μg/ml.

2.2. DNA Manipulations

The donor DNA (pCLPX-4) used for DNA transformation was constructed as follows; a *Sma*I/*Eco*RI fragment from pLP61 [18] carrying the *pilE* gene was cloned into pUC19. A promoter-less *cat* gene was then blunt-end cloned into the unique *Bsu*36 site immediately downstream of the *pilE* promoter. Consequently, transcription of the *cat* gene is under the control of the *pilE* promoter. For transformation experiments with donor DNAs containing 5' and 3' *pilE* flanking homologies plasmid pNG3005 was used, which contains a drug resistance marker (*ermC*) within *opaE* located immediately downstream of the *pilE* locus [20].

2.3. Transformation Protocols

Gonococci were lifted from plates on sterile Whatman paper fragments and the cells were re-suspended in 1 ml liquid growth medium containing 2 mM IPTG. The cells were then pre-incubated for 20 min to express RecA protein [21]. After 20 min pre-incubation, 1 μg of donor DNA was added to the culture and the cells were incubated for 3 hrs in a 5% CO_2 atmosphere. The cells were then pelleted by centrifugation and were plated on selective medium in the absence of IPTG. Following overnight growth, the individual drug resistant colonies were passed for an additional day on selective medium without IPTG before the total DNA was prepared. The *pilE* gene was then PCR amplified for DNA sequencing using the following primers: 5'-TCCCCTTTCAATTAGGAGT-3' and 5'-CCGATATATTATTTCCACC-3'.

Transformation frequencies were determined using the previously described single-colony transformation procedure [22]. The transformation frequencies were corrected for the initial piliation status of the recipent population as pilus minus bacteria are non-transformable. For transformation studies with *mutS* mutants, such corrections had a significant impact. Statistical analysis utilized Student's *t* test and reflect deviations from the standard mean. All other transformation experiments utilized the previously described plate-transformation protocol [23].

2.4. DNA Sequencing

DNA sequencing was performed either on an ABI DNA sequencer or using the manual dideoxy sequencing technique with the fragments being resolved on 8% polyacrylamide sequencing gels.

2.5. In Silico Analyses

Sequences were obtained from the NCBI database and aligned using the multiple sequence program MAFFT [24]. Thirteen *pilS* gene copies, that varied in size from 144 - 420 bp (mean = 340 bp), allowed a nucleotide alignment measuring 480 nucleotides in length (**Supplemental Table 1**). The alignment incorporated 366 gaps (due largely to the relatively small size of *pilS*6 copy 2 (144 bp) and *pilS*1 copy 6 (216 bp)) along with a few, short (~2 - 18 bp) regions that could not be aligned between all base pairs. This alignment was used to identify the sequence polymorphisms within *N. gonorrhoeae* MS11 *pilS* gene copies. A sequence logo was created from the constructed alignment in order to visualize the nucleotide divergence within the *pilS* alleles [25]. In order to determine the number of parsimony informative (each variant is present in at least two alleles) and singleton

(variant appears in only one sequence), the DNASP program was used [26]. Additionally, a script was created to identify singleton nucleotide divergences within the alignment. The PAML program was then used to find the average number of synonymous polymorphisms per synonymous site (Ks) and the average number of non-synonymous polymorphisms per non-synonymous site (Ka) [27]. DNASP was also used to perform the Tajima test in order to determine if non-neutral evolution is evident in any of the *pilS* alleles. The Poison Random Field model was used to detect if the nucleotide polymorphisms identified within the alignment were evolving under neutral expectations. To achieve this, the PRFMLE program was used to analyse the *pilS* alleles [28].

3. Results

3.1. In silico Analysis of Pil Sequence Heterogeneity

Previous analysis of *pilE* sequence divergence had revealed the presence of semi- and hyper-variant gene segments within the *pilE* gene [4]-[6]. In order to determine the total extent of *pil* gene sequence heterogeneity, the *pilS* gene copies were examined using various *in silico* tools. Thirteen *pilS* gene copies were aligned (**Table 1**; **supplemental data**) and used to determine the nucleotide polymorphisms as well as the intensity and direction of selection.

In order to visualize the regions containing prevalent nucleotide polymorphisms, the multiple sequence alignment of the 13 *pilS* copies was visualized as a sequence logo (**Figure 2**) [25]. Taller nucleotide stacks within the sequence logo denotes greater conservation, thus providing a visual represention of the relative frequency of the nucleotides in that position; areas with missing nucleotides correspond to regions where all, or most, sequences failed to align (*i.e.*, alignment gaps). As evident in **Figure 2**, *pilS* genes contain constant and

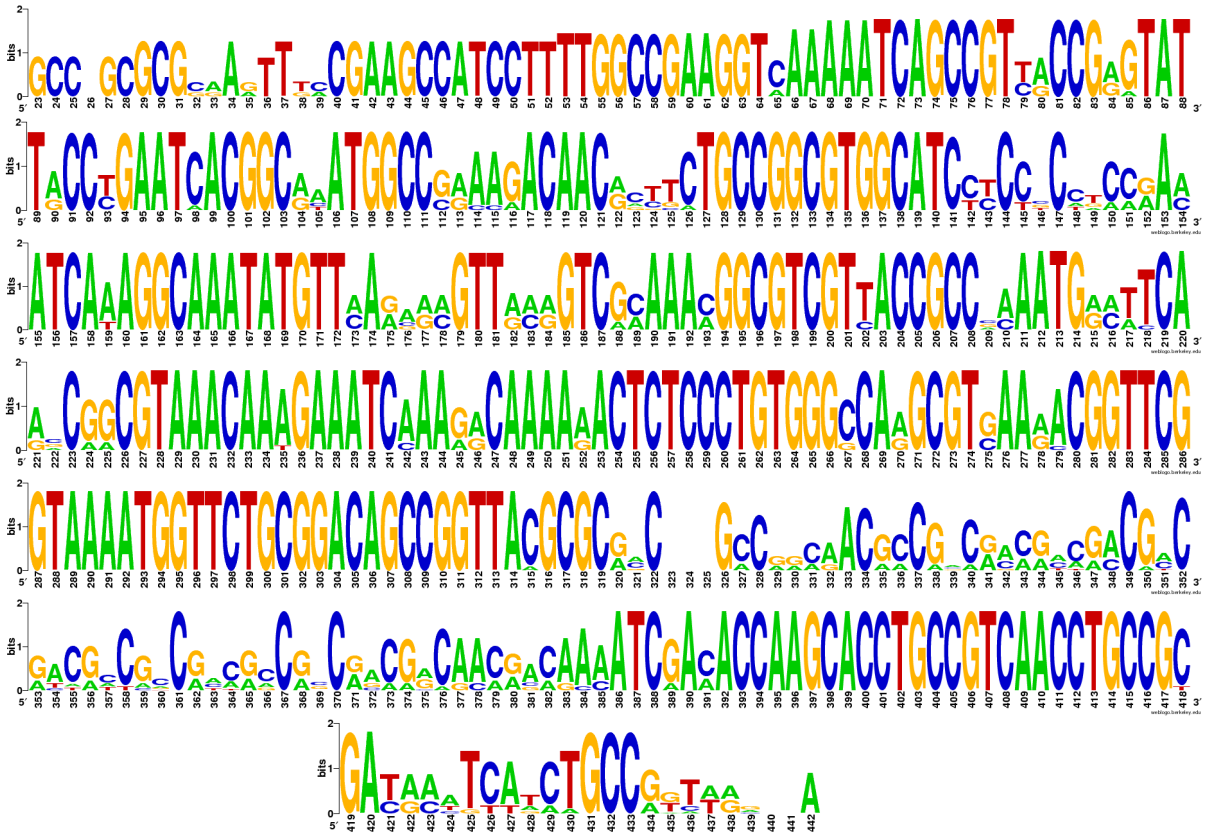

Figure 2. Sequence logo of *Neisseria gonorrhoeae* MS11 *pilS* A sequence alignment of the *pilS* gene copies from *Neisseria gonorrhoeae* strain MS11 (n = 13) from base pair 23 - 442 in the alignment [25]. The overall height of the nucleotide indicates the sequence conservation in the *pilS* alignment at that position. The height of nucleotides within a stack indicates the relative frequency of each nucleotide at that position. Areas missing nucleotides or that contain short stacks of nucleotides correspond to areas where the alignment fails to include all sequences (*i.e.*, alignment gaps).

Table 1. Summary of the descriptive statistics depicting the degree of polymorphism within each *pilS* allele.

		No. variants per site								
		Informative			Singleton					
pilS	No. Sites	2	3	4	2	3	4	K_s^*	K_a^t	K_a/K_s
All *pilS* (*n* = 13)	113	21	11	2	6	1	0	0.1247	0.1876	1.5042
*pilS*1 (*n* = 6)	200	18	9	0	24	0	0	0.0346	0.1583	4.5722
*pilS*2 (*n* = 2)	459	0	0	0	52	0	0	0.1591	0.1366	0.8587
*pilS*6 (*n* = 3)	357	0	0	0	30	3	0	0.0000	0.0946	99.0000
pilS except *pilS*1 copy 6 and *pilS*6 copy 2 (*n* = 11)	483	51	12	1	14	0	0	0.0444	0.1187	2.6714

Summary of the descriptive statistics depicting the degree of polymorphism within each pilS allele. Informative sites are parsimony informative, where each variant is present in at least two alleles. Singleton sites are where the variant occurs in only one sequence among the set [26] [27]. K_s^* constant represents the average number of synonymous polymorphisms per synonymous site (*i.e.*, those that would code for the same amino acid). K_a^t represents the average number of non-synonymous polymorphisms per non-synonymous site (*i.e.*, those that would code for a different amino acid). Informative sites are parsimony informative, where each variant is present in at least two alleles. Singleton sites are where the variant occurs in only one sequence among the set.

variable regions, with the the 5' end to the mid-gene region being relatively constant; the 3' region, however, contains an elevated degree of sequence diversity, especially between base pairs 320 - 385. As such, this initial analysis conforms to the notion of variable "mini-cassettes" interspersed with constant regions and reveals the so-called "hyper-variable" region towards the 3' end of the *pil* gene copies. This analysis also reveals several short stretches of non-identity towards the 5' end of the *pilS* gene copies.

To better envision the single nucleotide polymorphisms identified with the sequence logo algorithm, singleton base changes were extracted. A singleton base change is one that occurs in only a single sequence in the alignment. If the base change from the consensus sequence occurs in two or more *pilS* sequences, it is not considered a singleton base change and has not been included in **Figure 3**. Every *pilS* gene copy shows a single nucleotide polymorphism within the alignment except *pilS*1 copy 5. Therefore, this particular *pil* gene copy reflects the consensus sequence which suggests that *pilS*1 copy 5 shows similarity among the entire length of its sequence to the majority of the other *pilS* sequences. Most of the singleton changes occur between bases 320 - 385 (highlighted in grey) which is also observed within the sequence logo and corresponds to changes within the so-called "hyper-variable" region. Also evident in **Figure 3**, the 5' region also contains a number of singleton changes.

The *pilS* alignment was then analyzed to determine the degree of polymorphism between and within each *pilS* locus that contained more than two *pil* gene copies (*intra*-genic polymorphisms are excluded for *pilS*5 and *pilS*7). As can be seen in **Table 1**, the number of nucleotide changes that occur within two or more *pilS* sequences (referred to as parsimony-informative sites and designated as Informative) is indicated, along with the number of polymorphisms that are found only within a single *pilS* sequence (designated as Singleton); sites containing alignment gaps were excluded to ensure correct analysis (the number of bases analyzed are designated as No. Sites). Both the parsimony-informative and singleton statistics indicate the number of different nucleotides found at each site (2, 3 or 4). The constant K_s represents the average number of synonymous polymorphisms per synonymous site (*i.e.*, those that would code for the same amino acid), whereas the constant K_a represents the average number of non-synonymous polymorphisms per non-synonymous site (*i.e.*, those that would code for a different amino acid). Consequently, the more sequences within the alignment that contain a polymorphism at a designated synonymous or non-synonymous site will increase the K_s or K_a ratio, allowing the ratio of these two constants (K_a/K_s) to be used as an indicator of selection. A high ratio is indicative of positive selection. The data show that between all the *pilS* gene copies there is a large number of parsimony-informative sites, accounting for approximately 30% of all sites analyzed (2 = ~19%, 3 = ~10% and 4 = ~ 2%). When all 13 *pilS* gene copies were analyzed the K_a value was similar to the K_s value along the entire length of the alignment, indicating a lack of non-synonymous substitutions, which would account for relatively small changes in the resulting amino acid sequence upon unidirectional recombination with *pilE*. However, as *pilS*1 copy 6 and *pilS*6 copy 2 are much shorter than the other *pilS* copies and introduce large gaps at the 5' end of the alignment, analyses performed on the sequence alignment excluding these two *pilS* copies shows that there is a disparity between the amount of synonymous and non-synonymous polymorphisms, with the non-synonymous polymorphisms being more fre-

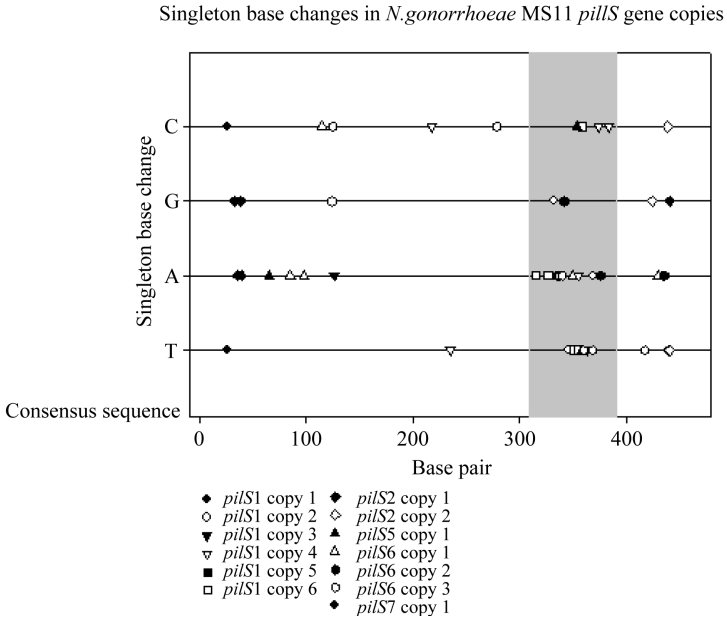

Figure 3. Singleton base changes in *N. gonorrhoeae* strain MS11 *pilS* gene copies. Graph of the entire *pilS* alignment showing divergence of certain *pilS* gene copies from the consensus sequence [26]. The segment corresponding to base pairs 320 - 385 of the alignment, highlighted in grey, shows the greatest divergence among all *pilS* alleles.

quent. This would result in a relatively high degree of sequence diversity in the resulting amino acid sequence upon unidirectional recombination with *pilE* (**Table 1**).

In order to visualize the number of synonymous and non-synonymous polymorphisms at different positions within the *pilS* genes, a sliding window approach was used to graphically represent the lack of uniformity. This was performed separately for synonymous and non-synonymous sites on all *pilS* alleles excluding *pilS*1 copy 6 and *pilS*6 copy 2. As the number of non-synonymous and synonymous sites is comparatively constant among the entire set of *pilS* genes, the same window sizes and sliding increments were used for both analyses. Interestingly, the 5' region of *pilS* showed a relatively high degree of synonymous substitutions when compared to the 3' end of the *pilS* alleles, with non-synonymous polymorphisms being more focused towards the 3' end of the genes (cf. **Figure 4(a)** and **Figure 4(b)**). This indicates that the sequence divergence seen with both the sequence logo (**Figure 2**) and the singleton analysis (**Figure 3**) towards the 5' end are enriched for synonymous base changes, while the sequences within the "hyper-variable" region (starting around base pair 310) contained a higher degree of non-synonymous substitutions.

In order to identify whether selection pressure was being applied near the 5' end of *pilS* sequences (with *pilS*1 copy 6 and *pilS*6 copy2 being excluded), the Tajima test was used to determine whether the nucleotide polymorphisms occurring within the *pilS* alignment are consistent with neutral expectations [26]. This test uses Tajima's D_T statistic, which is a scaled version of the ratio of the heterozygosity (k) of the population over the relative number of segregating sites based on a constant (S/a_1). Purifying selection is indicated if Tajima's D_T statistic is significantly less than zero, while a D_T value greater than zero is indicative of balancing selection. As seen in **Table 2**, the values of Tajima's D_T revealed evidence for balancing selection in the 5' region of the analyzed *pilS* alignment, while purifying selection appears to have taken place in the center of the so called "hyper- variable" region.

The Poisson Random Field model was also used to test if the polymorphisms are evolving under neutral expectations [28]. This test determines if polymorphisms occur under neutral expectations, at an estimated rate of $\mu > 0$ with a selective advantage of $\gamma > 0$, with a null hypothesis of $\gamma = 0$. When using this test on the *pilS* alleles, γ can be used to estimate the selective advantage and the nature of the selection pressure can be determined by using the sign (+ or −) of this value. When the Poisson Random Field model test was applied to the same *pilS* alignment excluding *pilS*1 copy 6 and *pilS*6 copy2, the analysis corroborated the Tajima test in that balancing

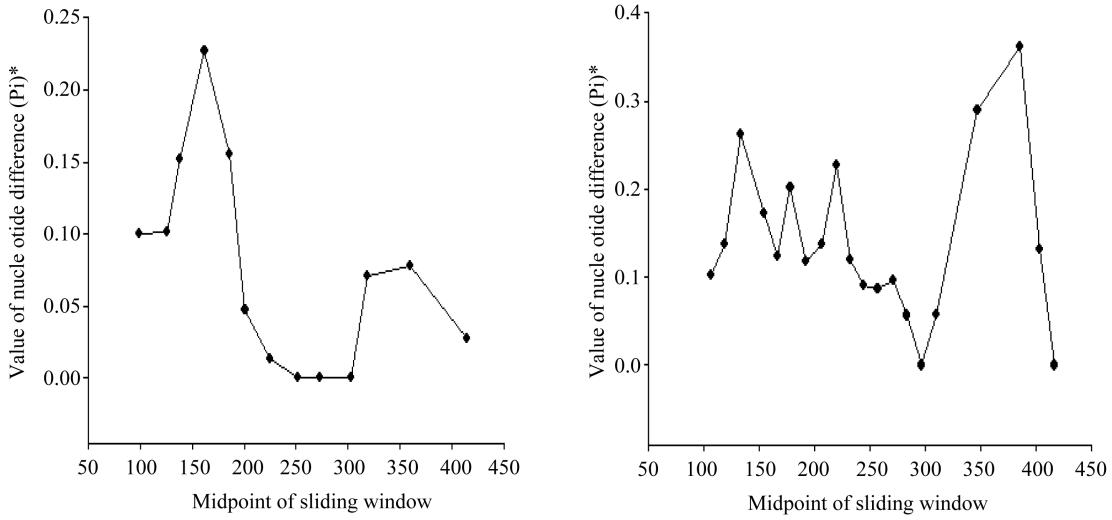

Figure 4. Synonymous and non-synonymous base changes in *pilS* gene copies excluding *pilS*1 copy 6 and *pilS*6 copy 2. Panel (A) Synonymous nucleotide substitutions (Pi) among *pilS* alleles based on a sliding window of 10 bp (moved in 5 bp increments). Panel (B) Nonsynonymous nucleotide substitutions (Pi) among *pilS* alleles based on a sliding window of 10 bp (moved in 5 bp increments). Each includes the Jukes-Cantor correction [27].

Table 2. Tajima's test on the entire set of aligned *pilS* alleles to determine any significant departure from neutral expectations.

Midpoint of *pilS* alignment (excluding *pilS*1 copy 6 and *pilS*6 copy2)	Tajima's D	Significance
172	1.8277	Indicates balancing selection $p < 0.10$
178	1.8277	Indicates balancing selection $p < 0.10$
181 - 184	1.9000	Indicates balancing selection $p < 0.10$
213	1.8277	Indicates balancing selection $p < 0.10$
347 - 349	−1.7085	Indicates purifying selection $p < 0.10$

Regions showing significant D_T values on the *pilS* alignment excluding *pilS*1 copy 6 and *pilS*6 copy 2 with a sliding window of 5 bp moved in 1 bp increments [26].

selection was indicated between base pairs 1 - 300 (**Table 3**). As there was only a short region undergoing purifying selection within the so-called "hyper-variable" region, this analysis was unable to detect any selection from base pairs 300 - 428 (not indicated in **Table 3**).

Overall, the above analyses were able to identify the constant and variable regions within the *pilS* gene copies; confirm that considerable sequence variation is present within the *pilS* gene copies; depending upon the position of the nucleotide polymorphism within the *pilS* gene copy, recombination with *pilE* may or may not lead to amino acid changes within the variant PilE polypeptides; and, the data also show that distinct regions within *pilS* are also evolving independently.

3.2. Pil Sequence Heterogeneity and the Role of the Mismatch Repair System

Given the variable nature of *pil* gene sequences, *pilE/pilS* recombination is a classic example of homeologous recombination, with *pil* recombination requiring the pairing and exchange of non-identical DNAs. An RNA-based assay was previously used to assess whether mismatch repair influences *intra*-cellular recombination between *pilE* and *pilS* in *N. gonorrhoeae* strain MS11 and showed that comparisons between isogenic *mutS+/MutS* bacteria, as well as isogenic $recB_{sup}mutS+/recB_{sup}mutS$ bacteria, equivalent signal intensities were observed, suggesting that the presence of an active mismatch repair does not appear to influence the efficiency of *intra*-cellular *pilE/pilS* recombination in strain MS11 [15]. Therefore, the mechanism for *pilE/pilS* recombination

Table 3. The poisson random field model assessing whether nucleotide polymorphisms within *pilS* are evolving under neutral expectations.

pilS (excluding *pilS*1 copy 6 and *pilS*6 copy 2)	γ (95% confidence interval)	Significance	Outcome
All *pilS* (*n* = 11)	2.77 ± 4.29	Statistically significant ($P = 3.49 \times 10^{-2}$)	Indicates balancing selection

The Poisson Random Field model was applied to detect if the polymorphisms within *pilS* are evolving under neutral expectations. This test was applied to the entire alignment of *pilS* along with each gene copy in the different loci. The Poisson Random Field model was able to detect balancing selection pressure on the polymorphisms within bp 1 - 300 of the *pilS* alignment when *pilS*1 copy 6 and *pilS*6 copy 2 were excluded [28].

appears to be "blind" to the inherent sequence non-identities.

We next examined whether *pil* sequence heterogeneity affected recombination when heteroduplex DNAs pair and recombine via the DNA transformation route. Transformation of recipients containing an IPTG-regulatable *recA* gene [21] with a defined *pilE*::*cat* construct allowed defined *pil/pil* recombinations to be established, with recombinant *pilE* genes being "locked-in" by plating transformant populations on plates lacking IPTG. DNA sequencing of *pilE* from chloramphenicol-resistant transformants allowed us to assess the extent of incorporation of the variant *pilE* gene sequence carried on the donor plasmid DNA into the recipient *pilE* gene. Following transformation of recipients expressing the *pilE* variant 7:30:2 allele with *Hinc*II-digested pCLPX-4 donor DNA, 20/21 recombinant *pilE*s that were sequenced (from several different transformation experiments) showed that incorporation of the plasmid donor DNA into the recipient *pilE* gene occurred until a 9 bp sequence non-identity was encountered between the two DNA molecules (**Figure 5**). The single exception was found to simply cross-in the drug resistance marker using regions of sequence identity that surrounds the promoter region. Therefore, the short sequence non-identities that were identified towards the 5' end of the sequence alignment logo (**Figure 2**) may be sufficient to impede recombination when non-identical *pil* DNAs recombine via the DNA transformation route.

In contrast to the observations presented above, when the same donor DNA was used to transform cells carrying a *mutS* mutation, in addition to the inducible-*recA* gene, two different types of *pilE* recombinant were recovered. These included those that simply crossed in the drug resistance marker and maintained the recipient *pilE* sequence and those that crossed in the marker as well as incorporating long tracts of the plasmid donor DNA, even across regions that contained multiple DNA sequence non-identities between the recombining DNAs (data not shown). Together, these observations suggest that with wild type bacteria, when non-identical *pil* DNAs recombine via the DNA transformation route, the recipient mismatch repair system will terminate the extent of recombination tracts at small regions of non-identity, whereas in the absence of mismatch repair homology constraints appear to be removed.

3.3. Sequence Heterogeneity, Mismatch Repair And Transformation Efficiency

In the absence of RecA protein (using the IPTG-regulatable strain), a *N. gonorrhoeae* strain MS11 *mutS* mutation elevated the phase transition rates within non-selected recipient populations (**Table 4**); in a *rec+/mutS* comparison, a 18.75X increase was observed, $p < 0.05$, $n = 10$; in a *recB* growth suppressor/*recB* growth suppressor *mutS* comparison, a 10X increase was observed, $p < 0.1$, $n = 10$. Although we did not explore this phenomenon further, pilus phase transitions in the absence of RecA are believed to be caused by frameshifting within the *pilC* locus [29].

We next examined whether heterologous *pilE/pilE* transformations are also more efficient in *mutS* mutants when the donor DNA carries a *pilE* gene with 5' and 3' *pilE* flanking sequences (pNG3005; [20]). The data presented in **Table 4** show that in a *rec+/mutS* comparison, there was a 15.6X increase in transformation efficiency with the *mutS* recipients (2.73×10^{-3} vs 42.6×10^{-3} for *rec+* and *mutS* respectively; $p < 0.05$, $n = 10$). Similarly, a 5.75X increase in transformation efficiency was observed in *recB* growth suppressor mutants that also carried a *mutS* mutation when compared to the *recB* suppressor parental strain (7.46×10^{-3} vs 42.9×10^{-3} for *recB* sup and *recB* sup *mutS* respectively; $p < 0.2$, $n = 10$). *pilE* sequence analysis of the *mutS* transformant populations also showed that *pilE* allelic exchange occurred in approximately 50% of those *pilE* genes that were sequenced ($n = 20$). Moreover, the sequencing data also indicated that intact *pilE* genes were exchanged without the incorporation of errors within the *mutS* transformant population. Despite this later observation indicating that transformant populations should be predominantly piliated, when these populations were scored microscopically for

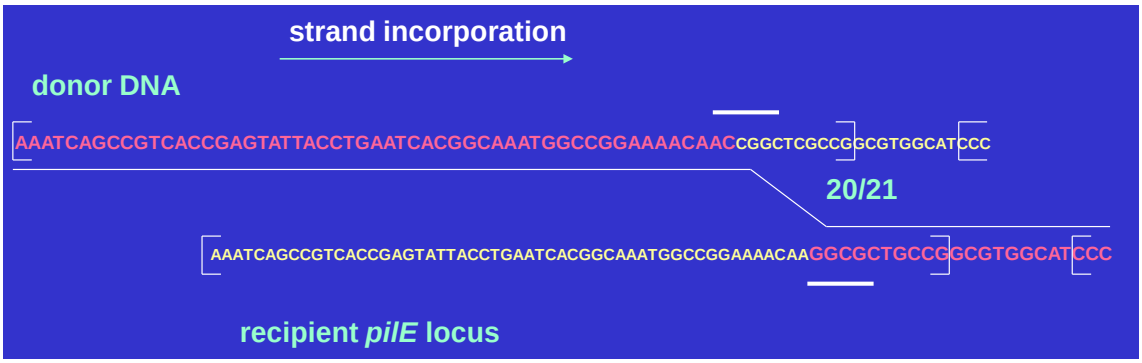

Figure 5. Effects of sequence heterogeneity following DNA transformation. Schematic representation indicating the extent of incorporation of the *Hinc*II-digested pCLPX-4 into a recipient *pilE* gene in wild type cells. The red highlighted sequence represents the recombinant sequence that was obtained following transformation. The white bars indicate the location of the sequence non-identities.

Table 4. Effect of *mutS* mutation DNA transformation rates and phenotypes.

	rec+	*mutS*	*recB*sup	*recB*sup *mutS*
Frequency[(a)]	2.73	42.6	7.46	42.9
Colony phenotype[(b)] recipient	99.9	88.1	95.7	83.1
transformant	32.2	20.5	27.4	24.4

[(a)]frequency = # transformants/μg donor DNA/cfu ($\times 10^{-3}$); [(b)]the colony phenotype (the percentage of piliated cells) of the recipient population was determined on cells plated at time 0 on plates without IPTG being present; the colony phenotype of the transformant population was determined on cells plated on selective medium. n = 10.

their piliation status, wild type *rec*+ bacteria showed a higher percentage of colonies that retained the piliated phenotype when compared to the isogenic *mutS* strain (32% vs 20%; **Table 4**).

4. Discussion

In this study, we examined the extent of *pil* sequence heterogeneity within *N. gonorrhoeae* strain MS11 and determined the impact of sequence divergence on various recombination profiles. *In silico* analysis revealed considerable sequence heterogeneity with non-synonymous nucleotide substitutions being primarily confined to the so-called "hyper-variable" region. Nonetheless, sequence heterogeneity was observed towards the 5' end of the *pilS* gene copies, however, they were predominantly synonymous substitutions which would conserve the primary amino acid sequence of the variant PilE polypeptides that are created during a *pilE/pilS* recombination event. Two statistical analyzes indicated that positive selective pressure is being applied on the 5' nucleotide polymorphisms in order to maintain synonymous substitutions. Consequently, the 5' and 3' ends of the *pilS* gene copies appear to be evolving independently. Despite the considerable observed sequence heterogeneity, and that *pilE/pilS* recombination involves recombination between these non-identical DNAs, we had previously found the presence or absence of an active mismatch repair system appeared to have little effect on the efficiency of *intra*-cellular gene conversion [15]. However, we did find that short sequence non-identities impeded incorporation of divergent donor DNA when delivered by the DNA transformation route, and that inactivating the mismatch repair system allowed longer tracts of divergent DNA to be incorporated during the transformation process. Consequently, the molecular mechanisms by which non-identical *pil* DNAs interact appears to dictate the recombination profile. Therefore, given the considerable sequence divergence within the *pil* system, this demands that any model for gene conversion be able to accommodate sequence heterogeneity.

A previous study on a different strain has also examined the effects of mismatch repair on *pilE* gene conversion and concluded that longer *pilS* gene tracts were incorporated into *pilE* recombinants in the absence of mismatch repair [29], which was also observed in our heteroallelic transformation experiments when MS11 *mutS* recipients were used. In the previous study, a kinetic assay, that scores the number of pilus minus outgrowths from a pilus plus colony over time, also indicated increased levels of antigenic variation in *mutS* mutants [29]. In contrast, we found little effect of a *mutS* mutation on the efficiency of *pilE/pilS* using a RNA based *intra*-cellular

assay [15]. A possible explanation for the discrepancies between the two studies could be that frameshifting within *pilC*, which is elevated with mismatch repair mutants [29], could lead to more pilus minus outgrowths in the kinetic assay, which are then scored as antigenic variants. Differences between the two studies were also observed on the impact of mismatch repair on transformation efficiency where we found that inactivation of mismatch repair caused elevated transformation frequencies. This difference could be due to, 1) our correction for the percentage of pilus minus bacteria in the recipient population as non-piliated gonococci do not engage in transformation; 2) our use of single colony transformations (each colony had between 10^5 - 10^6 cells), as a previous study had demonstrated that heavy recipient cell densities negatively impacted transformation frequencies [22]; and 3) differences in the donor DNA concentrations that were used in the two studies (2 ng vs 1 μg) [29]; this study].

The observation that inactivating the mismatch repair system can relieve sequence constraints on incorporation of non-identical DNAs may have some implications with regard to recombination that occurs during an infection. Mutator clones frequently arise through the accumulation of mutations within the mismatch repair system genes in an *in vivo* setting and have been proposed to drive adaptive mutation in bacteria [30]. Indeed, mutator clones apparently account for the emergence of epidemic *N. meningitidis* strains, which may facilitate quicker responses to changing environmental conditions [31] [32]. Therefore, the emergence of *Neisseria* mutator clones *in vivo* may allow some relaxation of the sequence constraints imposed by mismatch repair following horizontal transmission of chromosomal DNA via DNA transformation and could possibly expand the repertoire of *pil* genetic diversity. However, as a mismatch repair mutation significantly influenced pilus phase transition rates, the emergence of mutator clones *in vivo* is more likely to facilitate niche adaptation allowing an expansion of the disease profile (e.g., promote an invasive phenotype which is enhanced with pilus minus bacteria).

Acknowledgements

This work was supported in part by NIH grant 1R15 AI072720-01A1 to SAH.

References

[1] Meyer, T.F. and Hill, S.A. (2003) Genetic Variation in the Pathogenic *Neisseria* spp. In: Craig, A. and Scherf, A., Eds., *Antigenic Variation*, Academic Press, Waltham, 142-164.

[2] Swanson, J., Robbins, K., Barrera, O., Corwin, D., Boslego, J., Ciak, J., Blake, M. and Koomey, J.M. (1987) Gonococcal Pilin Variants in Experimental Gonorrhea. *Journal of Experimental Medicine*, **165**, 1344-1357. http://dx.doi.org/10.1084/jem.165.5.1344

[3] Meyer, T.F., Billyard, E., Haas, R., Storzbach, S. and So, M. (1984) Pilus Genes of *Neisseria gonorrhoeae*: Chromosomal Organization and DNA Sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 6110-6114. http://dx.doi.org/10.1073/pnas.81.19.6110

[4] Haas, R. and Meyer, T.F. (1986) The Repertoire of Silent Pilus Genes in *Neisseria gonorrhoeae*: Evidence for Gene Conversion. *Cell*, **44**, 107-115. http://dx.doi.org/10.1016/0092-8674(86)90489-7

[5] Haas, R., Veit, S. and Meyer, T.F. (1992) Silent Pilin Genes of *Neisseria gonorrhoeae* MS11 and the Occurance of Related Hypervariant Sequences among other Gonococcal Isolates. *Molecular Microbiology*, **6**, 197-208. http://dx.doi.org/10.1111/j.1365-2958.1992.tb02001.x

[6] Hagblom, P., Segal, E., Billyard, E. and So, M. (1985) Intragenic Recombination Leads to Pilus Antigenic Variation in *Neisseria gonorrhoeae*. *Nature*, **315**, 156-158. http://dx.doi.org/10.1038/315156a0

[7] Swanson, J., Morrison, S., Barrera, O. and Hill, S. (1990) Piliation Changes in Transformation-Defective Gonococci. *Journal of Experimental Medicine*, **171**, 2131-2139. http://dx.doi.org/10.1084/jem.171.6.2131

[8] Swanson, J., Berstrom, S., Robbins, K., Barrera, O., Corwin, D. and Koomey, J.M. (1986) Gene Conversion Involving the Pilin Structural Gene Correlates with Pilus+ to Pilus− Changes in *Neisseria gonorrhoeae*. *Cell*, **47**, 267-276. http://dx.doi.org/10.1016/0092-8674(86)90449-6

[9] Koomey, M., Gotschlich, E.C., Robbins, K., Bergstrom, S. and Swanson, J. **(**1987) Effects of *recA* Mutations on Pilus Antigenic Variation and Phase Transitions in *Neisseria gonorrhoeae*. *Genetics*, **117**, 391-398.

[10] Zhang, Q.Y., DeRyckere, D., Lauer, P. and Koomey, M. (1992) Gene Conversion in *Neisseria gonorrhoeae*: Evidence for Its Role in Pilus Antigenic Variation. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 5366-5370. http://dx.doi.org/10.1073/pnas.89.12.5366

[11] Cahoon, L.A. and Seifert, H.S. (2009) An Alternative DNA Structure Is Necessary for Pilin Antigenic Variation in *Neisseria gonorrhoeae*. *Science*, **32**, 764-767. http://dx.doi.org/10.1126/science.1175653

[12] Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. and Rehrauer, W.M. (1994) Biochemistry of Homologous Recombination in *Escherichia coli*. *Microbiological Reviews*, **58**, 401-465.

[13] Davidsen, T. and Tonjum, T. (2006) Meningococcal Genome Dynamics. *Nature Reviews Microbiology*, **4**, 11-22. http://dx.doi.org/10.1038/nrmicro1324

[14] Davidsen, T., Tuven, H.K., Bjoras, M., Rodland, E.A. and Tonjum, T. (2007) Genetic Interactions of DNA Repair Pathways in the Pathogen *Neisseria meningitidis. Journal of Bacteriology*, **189**, 5728-5737. http://dx.doi.org/10.1128/JB.00161-07

[15] Hill, S.A. and Davies, J.K. (2009) Pilin Gene Variation in *Neisseria gonorrhoeae*: Reassessing the Old Paradigms. *FEMS Microbiological Reviews*, **33**, 521-531. http://dx.doi.org/10.1111/j.1574-6976.2009.00171.x

[16] Swanson, J. (1982) Colony Opacity and Protein II Compositions of Gonococci. *Infection and Immunity*, **37**, 359-368.

[17] Bergstrom, S., Robbins, K., Koomey, J.M. and Swanson, J. (1986) Piliation Control Mechanisms in *Neisseria gonorrhoeae. Proceedings of the National Academy of Sciences of the United States of America*, **83**, 3890-3894. http://dx.doi.org/10.1073/pnas.83.11.3890

[18] Hill, S.A., Morrison, S.G. and Swanson, J. (1992) The Role of Direct Oligonucleotide Repeats in Gonococcal Pilin Gene Variation. *Molecular Microbiology*, **4**, 1341-1352. http://dx.doi.org/10.1111/j.1365-2958.1990.tb00713.x

[19] Hill, S.A., Woodward, T., Reger, A., Baker, R. and Dinse, T. (2007) Role for the RecBCD Recombination Pathway for *pilE* Gene Variation in Repair Proficient *Neisseria gonorrhoeae*. *Journal of Bacteriology*, **189**, 7983-7990. http://dx.doi.org/10.1128/JB.00980-07

[20] Wainwright, L.A., Pritchard, K.H. and Seifert, H.S. (1994) A Conserved DNA Sequence Is Required for Efficient Gonococcal Pilin Antigenic Variation. *Molecular Microbiology*, **13**, 75-87. http://dx.doi.org/10.1111/j.1365-2958.1994.tb00403.x

[21] Seifert, H.S. (1997) Insertionally Inactivated and Inducible *recA* Alleles for Use in *Neisseria*. *Gene*, **188**, 215-220. http://dx.doi.org/10.1016/S0378-1119(96)00810-4

[22] Hill, S.A. (2000) Opa Expression Correlates with Elevated Transformation Rates in *Neisseria gonorrhoeae*. *Journal of Bacteriology*, **182**, 171-178. http://dx.doi.org/10.1128/JB.182.1.171-178.2000

[23] Gunn, J.S. and Stein, D.C. (1996) Use of a Non-Selective Transformation Technique to Construct a Multiply Restriction/Modification-Deficient Mutant of *Neisseria gonorrhoeae*. *Molecular and General Genetics*, **251**, 509-517.

[24] Katoh, S. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772-780. http://dx.doi.org/10.1093/molbev/mst010

[25] Crooks, G.E, Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Research*, **14**, 1188-1190. http://dx.doi.org/10.1101/gr.849004

[26] Hartl, D.L., Moriyama, E.N. and Sawyer, S.A. (1994) Selection Intensity for Codon Bias. *Genetics*, **138**, 227-234.

[27] Librado, P.R.J. (2009) DnaSP v5: A Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics*, **25**, 1451-1452. http://dx.doi.org/10.1093/bioinformatics/btp187

[28] Yang, Z. (2007) PAML 4: A Program Package for Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, **24**, 1586-1591. http://dx.doi.org/10.1093/molbev/msm088

[29] Criss, A.K., Bonney, K.M., Chang, R.A., Duffin, P.M., LeCuyer, B.E. and Seifert, H.S. (2010) Mismatch Correction Modulates Mutation Frequency and Pilus Phase and Antigenic Variation in *Neisseria gonorrhoeae. Journal of Bacteriology*, **192**, 316-325. http://dx.doi.org/10.1128/JB.01228-09

[30] Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H. and Godelle, B. (1997) Role of Mutator Alleles in Adaptive Evolution. *Nature*, **387**, 700-702. http://dx.doi.org/10.1038/42696

[31] Richardson, A.R. and Stojiljkovic, I. (2001) Mismatch Repair and the Regulation of Phase Variation in *Neisseria meningitidis*. *Molecular Microbiology*, **40**, 645-655. http://dx.doi.org/10.1046/j.1365-2958.2001.02408.x

[32] Richardson, A.R., Yu, Z., Popovic, T. and Stojiljkovic, I. (2002) Mutator Clones of *neisseria meningitidis* in Epidemic Serogroup A Disease. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6103-6107. http://dx.doi.org/10.1073/pnas.092568699

Supplemental Table 1. List and length of the 13 *pilS* sequences used in this study.

pil	Sequence length (bp)
*pilS*1 copy 6	216
*pilS*1 copy 5	345
*pilS*1 copy 4	348
*pilS*1 copy 3	363
*pilS*1 copy 2	405
*pilS*1 copy 1	363
*pilS*2 copy 1	420
*pilS*2 copy 2	405
*pilS*5 copy 1	357
*pilS*6 copy 1	351
*pilS*6 copy 2	144
*pilS*6 copy 3	306
*pilS*7 copy 1	396