

## Molecular characterization of repetitive DNA sequences from a B chromosome

Danna G. Eickbush, Thomas H. Eickbush, and John H. Werren

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received December 27, 1991 / in revised form March 17, 1992

Accepted April 2, 1992 by H.F. Willard

**Abstract.** In the parasitic wasp *Nasonia vitripennis*, certain males carry a B chromosome, called PSR (paternal sex ratio), which causes the compaction and subsequent loss of the paternal chromosomes in fertilized eggs. Because *Nasonia* are haplo-diploid, this leads to the production of all-male broods. Three families (PSR2, PSR18, PSR22) of related, tandemly repetitive DNAs were shown to be present solely on the PSR chromosome. These three families shared two conserved, palindromic DNA sequences, which may play a role in either PSR function or amplification of the tandem arrays. The tandem repeat family NV79 was determined to be present on the PSR chromosome as well as on at least one of the A chromosomes. This shared repeat as well as two repeat families (NV85, NV126) that were localized on the A chromosomes were detected in two sibling species of *N. vitripennis*. NV79 and NV126 were also found in the more distantly related species, *Trichomalopsis dubius*.

### Introduction

B chromosomes are non-vital supernumerary chromosomes that have been found in hundreds of plant and animal species (Jones and Rees 1982). Many B chromosomes have been shown to be transmitted at higher rates than expected from Mendelian inheritance and, therefore, possess a "drive". Nevertheless, little is known about the origin, replication drive, or molecular structure of B chromosomes.

PSR (paternal sex ratio) is a small B chromosome that occurs in some populations of the parasitoid wasp, *Nasonia vitripennis*. Only males carry this chromosome. The normal mode of sex determination in *Nasonia*, as in other hymenoptera, is haplo-diploidy. Females control the sex ratio of their offspring by altering the percentage of eggs that are fertilized; males develop from unferti-

lized eggs while females develop from fertilized eggs. Although typically females produce a high proportion of daughters, females that mate with males carrying PSR produce all male progeny (Werren et al. 1981; Werren and Assem 1986). Genetic analysis in conjunction with cytological observations of fertilized eggs reveals that during the first mitotic division the PSR chromosome causes the paternal A chromosomes to condense into a compact mass, which is subsequently lost (Werren et al. 1987). The PSR chromosome is itself insensitive to the effect and is passed on to the next generation (Nur et al. 1988; Werren 1991). Because it survives at the expense of the entire paternal genome, the PSR chromosome can be considered an extreme example of selfish DNA.

PSR is the only known B chromosome that accomplishes its replication drive by actively eliminating the paternal chromosomes. Characterizing the molecular structure of the PSR chromosome is a first step toward understanding the genetic and molecular basis of this phenomenon. In addition, it will provide data on the molecular structure of B chromosomes, of which little is currently available. Finally, understanding the mechanism by which PSR causes the compaction of the paternal genome may lead to unique insights into the normal process of chromosome condensation and/or mitosis.

Initially, genomic clones were isolated that contained PSR specific DNA sequences. Analysis of one such clone revealed a 171 by tandem repeat (PSR2) found on the PSR B chromosome but not on the five normal A chromosomes (Nur et al. 1988). In this report, we describe two additional families of tandem repetitive DNA that were specific to the PSR chromosome. A fourth repeat was identified that was enriched on the PSR chromosome but was also found in the genome of *N. vitripennis* and several related species.

### Materials and methods

*Isolation of PSR specific sequences.* Construction of the *N. vitripennis* genomic library and generation of the PSR specific phage sub-

library were as described by Nur et al. (1988). The sublibrary was screened by placing 75  $\mu$ l aliquots of the phage stocks in individual wells of a 96 well microtiter plate. Small aliquots of the lysate were spotted onto bacterial lawns (KH802) with a replica plater. After growth at 37° C for 16 h, the resultant plaques were screened by the method of Benton and Davis (1977). Phage DNA was purified as described by Maniatis et al. (1982).

*Restriction digests and Southern blots.* Phage and genomic DNA were digested with excess amounts of the indicated enzymes under the conditions described by the supplier. DNA was subjected to electrophoresis on horizontal 1 % agarose gels and transferred to nitrocellulose after partial depurination using standard conditions (Maniatis et al. 1982).

*Hybridizations.* Individual DNA fragments to be used as probes were cut out of 1 % low gelling temperature agarose gels, heated to 65° C, extracted twice with warm phenol (37° C, extracted three times with ether, and the DNA ethanol precipitated. Fragments were labeled with  $\alpha$ -<sup>32</sup>P dATP according to the procedure provided with the Amersham random labeling kit.

Hybridization of radioactively labeled probe to the DNA blot was conducted in 4 x SSC (0.6 M NaCl, 0.06 M sodium citrate), 5 x Denhardt's (0.1 % bovine serum albumin, 0.1 % polyvinylpyrrolidone, 0.1 % Ficoll), 25 mM sodium phosphate, 0.1 % sodium pyrophosphate, 0.1 % SDS, 250  $\mu$ g/ml denatured calf thymus DNA, 10% dextran sulfate (omitted in the low stringency hybridizations). High stringency hybridizations were done at 75° C and the final wash was in 0.1 x SSC, 0.1 % SDS. Moderate stringency hybridizations were done at 65° C and the final wash was as above. Low stringency hybridizations were done at 60° C and the final wash was in 1 x SSC, 0.1 % SDS. All washes were conducted at the same temperature as the hybridization.

*Sequence analysis.* Fragments containing the desired repeats were cloned into the plasmid vector puc13 or one of its derivatives (puc18 or puc19). Sequencing was done using the double stranded dideoxy nucleotide method (Sanger et al. 1977) with a US Biochemical Corporation kit. Regions of nucleotide similarity were initially localized using the matrix analysis programs available in the Mac-Vector Analysis Software (Biotechnologies). Specific sequences were submitted to the EMBL Data Library and were assigned the following accession numbers: nv104-6 (X64089), nv126-6 (X64090), nv79-16a (X64091), nv85-7 (X64092), psr105-3 (X64093), psr18-1 (X64094), psr2-1 (X64095), psr22-2 (X64096).

*Nomenclature.* Throughout the text, the abbreviation PSR refers to the entire B chromosome; PSR2 (18, 22) refers to a family of repeats that cross hybridize under moderate stringency criteria; ~psr# refers to a specific lambda phage clone; psr#-1 refers to a puc subclone from the lambda phage of the same number.

## Results

### *Characterization of PSR specific sequences*

Initial studies were directed toward isolating DNA sequences specific to or enriched on the PSR chromosome. A lambda genomic library constructed from total DNA isolated from males carrying PSR was screened in parallel with labeled genomic PSR DNA or genomic wild-type DNA. Fifty-two clones hybridized to the PSR probe but not to the wild-type probe. These clones were selected as containing PSR specific sequences and together comprised the PSR specific sublibrary. The sublibrary was then screened with labeled DNA insert isolated from

randomly selected PSR specific phage. The 52 clones were divided into three families based upon cross hybridization revealed during these screenings. These families were PSR2 (28 clones), PSR18 (17 clones), and PSR22 (7 clones). With the exception of 4 clones (discussed below), members from different families did not cross hybridize under the moderate stringency criteria used. Subsequent experiments revealed that increasing the hybridization temperature to 75° C divided the PSR18 family into two subfamilies that no longer cross hybridized. These two subfamilies are represented by ~psr18 and )psr105 in the text and figures. The PSR2 and PSR22 families could not be further subdivided at the higher temperature (L. Beukeboom, personal communication).

To determine what fraction of the DNA inserts from the PSR specific clones contained repetitive DNA, the DNA from 27 lambda clones was isolated for restriction enzyme analysis. Of these clones 11 belonged to the PSR2 family, another 11 belonged to the PSR18 family, and 5 were members of the PSR22 family. Restriction enzyme mapping was consistent with the conclusion that these clones comprised long arrays of short, tandemly repeated units. This conclusion was based primarily on the finding that for 24 clones a restriction enzyme could be found that cleaved the entire insert into a single fragment of approximately 200 bp or to a ladder array consisting of multiples of this monomer size. All clones within the same family were cleaved by a common restriction enzyme. Because the size of the lambda inserts ranged from 10 to 15 kb, approximately 50-75 of these repeats were assumed to be tandemly arrayed in each clone. The organization of the remaining 3 clones (2 from the PSR18 family as well as ~psr22) was not known because a restriction enzyme could not be found that digested the DNA into the distinctive pattern observed for the other clones. Southern analysis suggested that these 3 clones were composed entirely of repetitive DNA (data not shown). Presumably, they contained arrays of the repeats, the majority of which had lost the enzyme cleavage site used to recognize the repeat family. No non-repetitive DNA was detected in the 27 clones analyzed.

Southern analysis of genomic DNA confirmed that the majority of sequences comprising the PSR2, PSR18 and PSR22 families were present on the PSR chromosome in long, tandem arrays. When total genomic PSR DNA was cut with the restriction enzyme previously determined to cleave within the repeat family and then probed with a member of the family, a ladder consisting of multiples of the monomer length was observed (Fig. 1, lanes 1, 7, and 11). On the other hand, if a second enzyme that had been determined not to cut any of the clone inserts was used most of the hybridization was confined to fragments greater than 20 kb for the PSR2 and PSR18 families and greater than 10 kb for the PSR22 family (Fig. 1, lanes 3, 5, and 9). Hybridization to minor bands was thought to represent interspersions of the repeats with other DNA sequences, or more likely repeats within the array that contained restriction site polymorphisms. The same strain of *Nasonia* (Lab 11) minus the PSR chromosome was also tested for the presence of these repeats. The lack of hybridization to the

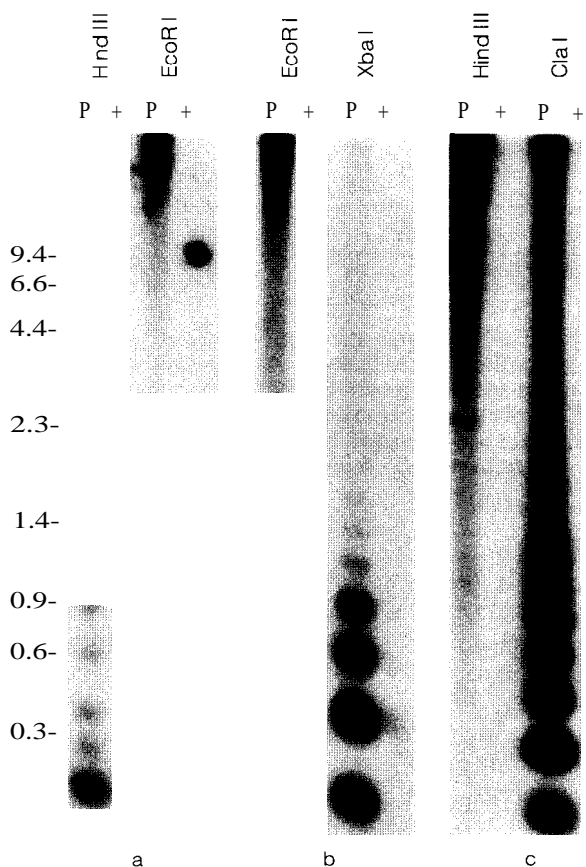


Fig. 1 a-c. Genomic blots of DNA from paternal sex ratio (PSR) (P) and wild-type (+) males probed with either a PSR2, b PSR18, or c PSR22. For each lane, 3  $\mu$ g of DNA was digested with the indicated restriction enzyme. Hybridizations were conducted under the moderate stringency criteria described in Materials and methods. The PSR2 probe was the entire insert isolated from  $\sim$ psr%; the PSR18 probe was the entire insert isolated from  $\sim$ psr105; the PSR22 probe was the entire clone  $\sim$ psr22. DNA standards, in kilobases, are given at the left

wild-type (Lab II) genomic DNA (Fig. 1, even lanes) indicated that the PSR2, PSR18, and PSR22 families were essentially absent on the A chromosomes. (A few copies may have gone undetected; see below.) Therefore, the majority of these repeats were confined to the PSR chromosome and each family was localized to a single (or perhaps a few) long array interspersed with few, if any, other sequences.

#### Sequence analysis of PSR specific repeats

Individual repeats from each family were subcloned and sequenced. A summary of the sequence data is presented in Fig. 2. For the PSR2 family, five repeats were sequenced from four different lambda clones ( $\sim$ psr2, Apsr12, Apsr96, Apsr127). All five repeats were 171 by in length. Pairwise comparison revealed 92%-99% sequence identity. For the PSR22 family, three repeats were sequenced from Apsr22. One repeat (1a) had a single base pair insertion and a single base pair deletion

relative to the other two repeats thus all three repeats were 183 by in length. The three repeats showed 90%-95% sequence identity. Within the PSR18 family, six repeats were subcloned and sequenced from four different lambda clones. Based on the hybridization data described in the previous section,  $\sim$ psr105 and  $\sim$ psr10 belonged to the PSR105 subfamily while Apsr18 and Apsr13 belonged to the PSR18 subfamily. Comparison of these sequences revealed a greater diversity in terms of repeat length than seen for either the PSR2 or PSR22 families. The sequenced PSR18 repeats ranged in length from 154 by (psr18-7) to 214 by (psr105-3). Owing to these differences in repeat lengths, only those sites (146 by total) present in all six repeats were used to calculate the percentage sequence identity. The level of identity between the two clones in the PSR105 subfamily was 99%. The level of sequence identity among the PSR18 subfamily repeats varied from 86% to 95% with an average identity of 88%. Among members of the different subfamilies, the sequence identity varied from 79% to 86% with an average identity of 83%.

The next step undertaken was a search for sequence identity among the different repeat families. Each sequenced repeat was compared with the other known repeats by matrix analysis. Compilation of the results indicated that two DNA regions were similar in members of all three families. Four repeats representing the PSR specific families and subfamilies have been aligned in Fig. 3. The percentage sequence identity along the length of the PSR2 and PSR18 (PSR105) repeats was 73% (71%). [Only those sites present in all 4 repeats (161 by total) were used in these calculations.] This alignment clearly suggested that the PSR2 and PSR18 families were related and probably arose from a common ancestor. The total sequence identity between these two families and PSR22 was significantly lower (45% for PSR22 versus PSR2 or PSR18; 50% for PSR22 versus PSR105).

The two conserved regions (boxed in Fig. 3) suggested that PSR22 was related to the PSR2 and PSR18 families. These two conserved regions were more striking when it was observed that they both harbor palindromic sequences. In the case of region 1, two of the four repeats contained a perfect 14 by palindrome while in region II, three out of the four repeats had a perfect 10 by palindrome (Fig. 3). Although the two conserved regions were originally detected by computer analysis, it was reasonable to ask whether they were significantly more conserved than other regions of the repeats. This question can be approached in several different ways. Results of one method are shown in Fig. 4. Only the 161 by sites present in all four repeat families (i.e. deletions/insertions were excluded) were used to search the four repeat families for highly conserved stretches. At each site, paired comparisons were made among the four repeats. This gave a total of six comparisons per site (i.e. 18 to 105, 18 to 2, 18 to 22, 105 to 2, 105 to 22 and 2 to 22). Each site was assigned a difference score based upon the paired comparisons; for example, a score of zero occurred when all four sequences had the same nucleotide at a site, a score of three occurred when one sequence was different from the other three, and a score

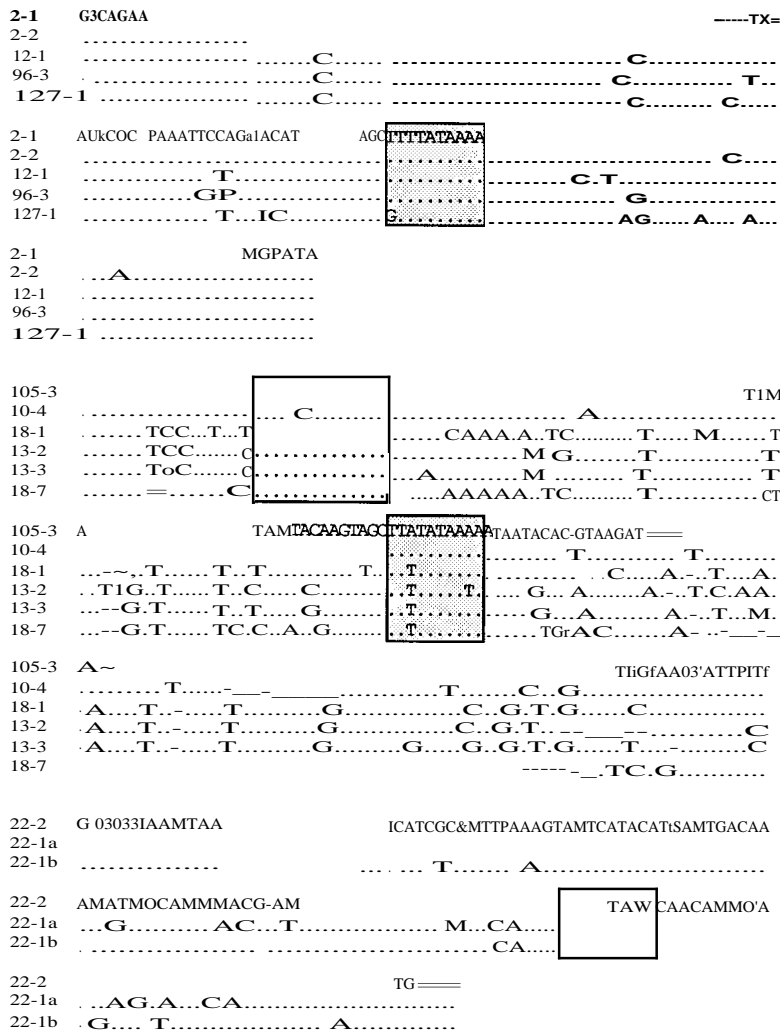


Fig. 2. Nucleotide sequences of the 14 cloned PSR repeats. The repeats have been grouped according to their family membership: five PSR2 repeats from four lambda clones; six PSR18 repeats from four lambda clones; and three PSR22 repeats from a single lambda clone (psr22-1a and psr22-1b are consecutive repeats). Letters denote the base changes from the top sequence while dashes indicate deletions. Boxed segments correspond to the two conserved regions described in Fig. 3

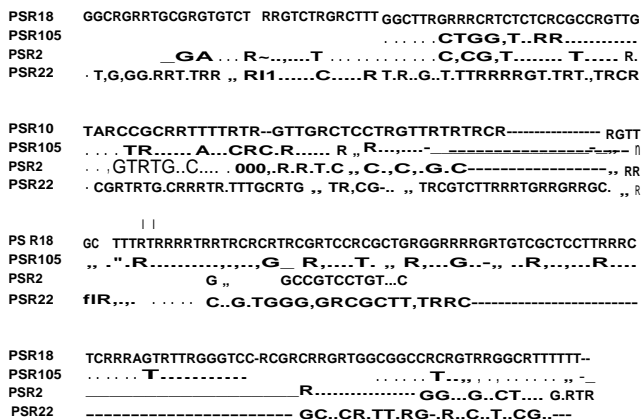


Fig. 3. Sequence comparison of four repeats representing the different PSR specific families and subfamilies. PSR2 corresponds to psr2-1; PSR18 to psr18-1; PSR105 to psr105-3; PSR22 to psr22-1b. Letters indicate base changes from the psr18-1 repeat and dashes correspond to deleted nucleotides. Boxed sequences are the conserved, palindromic regions I (white) and II (stippled)

of six occurred when all four sequences were different at a site. The frequency distribution of difference scores for the 161 base sites was 36% (0), 40% (3), 10% (4), 13% (5), 1% (6).

To detect stretches along the repeat that were relatively more conserved, the regions flanking each base site were compared. This was accomplished by creating a "window" of 11 by (5 to each side of the base site in question) and calculating a summed difference score (Kreitman and Hudson 1991). A window of 11 was chosen because it was large enough to develop a statistic with reasonable resolving power and yet small enough to detect short stretches of high sequence conservation. Within any window, the hypothetical maximum difference score was 66 and the minimum was 0. Regions of high conservation have low difference scores. In Fig. 4 it can be seen that the summed difference scores ranged from a high of 48 in variable regions to a low of 5 in the palindromic region I and 6 in palindromic region II. Given the frequency distribution of site difference scores above, the probability of getting a particular summed difference score in a stretch of 11 bases can be calculated using multinomial sampling theory. Using

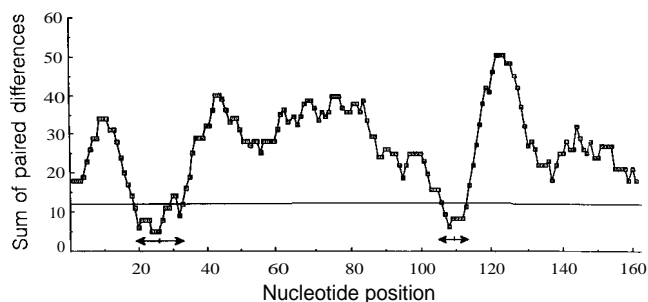


Fig. 4. Plot of the summed difference scores as a function of nucleotide position with the PSR repeat. The 161 nucleotide sites (i.e. deletions were excluded) shared among the four repeat families were used. Arrows represent the conserved palindromic regions I and II. As indicated by the solid horizontal line, two regions with scores of 11 or less are significantly conserved at the  $P < 0.001$  level

this approach, the probability of getting a summed difference score of 5 or less is  $2.3 \times 10^{-4}$ . The probability of getting a difference score of 11 or less is  $1.02 \times 10^{-2}$ . Therefore, 11 served as the  $P < 0.01$  significance level shown in Fig. 4.

The values above are the probability of getting a particular summed difference score by chance in a single stretch of 11 bp. There are 161 such stretches and of these, 19 have summed difference scores of 11 or less. However, the 161 stretches overlap partially and therefore were not independent of each other. This lack of independence made it more difficult to determine the probability of getting 19 such scores out of 161 samples. This was determined by the "bootstrap" method. Randomly sampling from the frequency distribution of the base site difference scores above, 600 different stretches of 161 base site difference scores were generated. Based upon these, the chance of getting a 161 by stretch containing 19 windows with summed difference scores of 11 or less is  $P < 0.002$ . We concluded that the two regions were significantly conserved.

#### Search for PSR repeats in related species

No one has succeeded in tracing the origin or documenting the ancestry of a B chromosome. Similarly, the origin of the PSR chromosome is unknown. Two alternatives are likely: (1) the PSR chromosome arose from one of the five A chromosomes in *N. vitripennis* with the subsequent expansion of the repeats, or (2) it was derived from chromosomal fragmentation during an interspecies hybridization between *N. vitripennis* and one of its two sibling species - *N. longicornis* or *N. giraulti* (Darling and Werren 1990). *N. longicornis* and *N. giraulti* also have five A chromosomes. There are several lines of circumstantial evidence to suggest the second possibility. First, microorganism induced cytoplasmic incompatibility between *N. vitripennis* and its sibling species is known to result in fragmentation of the paternal chromosomes in interspecies crosses (Breeuwer and Werren 1990). Second, it is known that cytoplasmic incompatibility occa-

sionally gives rise to centromere containing chromosomal fragments that in general are stably inherited through males (Ryan et al. 1985, 1987; L. Beukeboom and J. Werren, in press). And third, opportunities for interspecies mating occur in natural populations (Darling and Werren 1990).

Because the PSR chromosome may have arisen from one of the sibling *Nasonia* species, it was of interest to determine whether or not the PSR repeats were present in these two species. When Southern blots of genomic DNA were probed with a PSR2 repeat under low stringency criteria, weak hybridization was observed to total DNA from the two sibling species (Fig. 5). Under these conditions, weak hybridization was also observed to genomic DNA from wild-type *N. vitripennis*. The lack of detectable bands in some lanes (Nv, lanes B, C, D and Ng, lanes C, D) is most likely because the hybridization is distributed over several fragments. However, the intensity of hybridization and size of the EcoRI fragments (A lanes) were similar in all three species. Low molecular weight fragments in *N. longicornis* and *N. giraulti* were observed. The size of these small fragments (200 and 400 bp) suggested a short, tandem repeat. These fragments may represent that portion of the A chromosome from which the PSR chromosome was derived. These sequences would have subsequently been amplified on the PSR chromosome. Although this result did not dis-

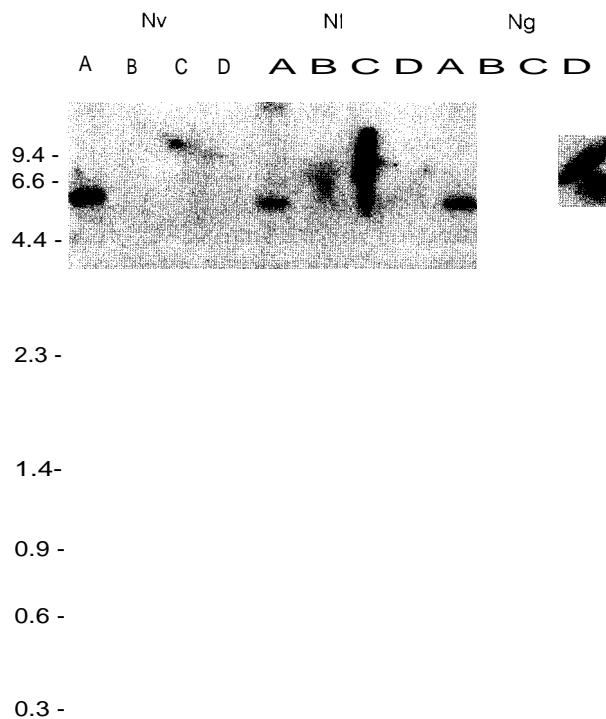


Fig. 5. Genomic blot of DNA from *Nasonia vitripennis* (Nv), *N. longicornis* (NI), and *N. giraulti* (Ng) probed with PSR2. Each lane contains 3  $\mu$ g of DNA digested as follows: A, EcoRI; B, HindIII; C, HaeIII; D, AluI. Hybridization was conducted under low stringency conditions. The PSR2 probe was the entire insert isolated from ~psr12. DNA standards are given in kilobases

```

nv79-16a  TCGRGTTCGRRTCGTCRCGRTRTCGGRGGRAGGCGRRRCGRGGCGRRRTTTRTTTARRRCRATRARCCCTTARRCCCRCCCTRCCTRTAGRCR
nv79-16b  R.....T.....G...R.....T.....G.....R.G..C.T.....C...T.....C...C...
nv79-14   R.C.R.G.....R..G.C.A.....R.O.....G.....G.....R.....R..C.....C...C...
nv79-15   R.....G...R.....T...G.....C.....R..T...G.....C.....C...

```

```

nv85-7   TRRGTTCTCGGCTTRCCCRCTRTTRTCCCTTCTRTTGCGRGTGGTTT-RTTCTCTRGTRGTRRRRRRTRTTRCRGCRGGTTG
nv85-9   .....T.....T.....G.R.....AT.....
nv104-6  .....R.....T,RGRR.....T.T.RRGRC.....T...GT...6.....C..R.....C...T.....

```

```

nv85-7   GTTGRARTCGRRGGTTGGRRRTCTGGRCRRCTGGGCGCTGRTGTTGTCRRRRGRRCTGRRTCRCGRRTTRGRTCTTTGTCGR
nv85-9   .....R.....T.....T..T.....RR.....T.....TX.....G.....
nvIG4-6  .....R.....T.....GR.....C..TR.GT.GR.....C.A.G.RG..G..R.TC.CG.....C...

```

```

nv126-6  CTGGGTTTRHIGCRCTTGGTTTTTRGRGTTTTRCCTCGCTTGGCTCGCCTTRCGCC
nv126-1a  C.....T.G...T..T..RT..TCT..G.C...R..R.C...
nv126-1b  R.....T.....T.....T..TCT.....C.T..

```

```

nv126-6  GRTGTGTCGRRRGTTTTRTTRTTRTTRCTGRTCTCGGGGGCGRTGGTGTGR
nv126-1a  .....O.....A.....T..A.....G.....
nv126-1b  .....A.....T..R...A.....G.....

```

tinguish between the two possible origins discussed above, it raised the possibility that the PSR chromosome could have arisen from a region that has been conserved in the three sibling species. Experiments are in progress to clone these sequences from an *N. longicornis* genomic lambda library in order to determine their level of sequence identity to the PSR2 family.

#### Characterization of a shared repeat

One repeat family was identified that appeared to be enriched in the PSR genome but was also present on the A chromosomes of wild-type males. nv79 was originally isolated as a PSR specific clone. However, a preliminary Southern blotting analysis revealed hybridization to genomic DNA isolated from wild-type males at an intensity that was approximately sevenfold lower than observed to PSR genomic DNA. This clone did not cross hybridize with any of the PSR specific families previously described. Restriction enzyme analysis suggested that the insert was composed of tandemly repetitive DNA since 11 restriction enzymes having hexanucleotide recognition sequences (EcoRI, HindIII, XbaI, XhoI, BamHI, KpnI, SstI, PstI, PvuI, PvuII, BglII) and two having tetranucleotide recognition sequences (AluI, HpaII) did not cleave the 13 kb insert. Although several restriction enzymes did cleave the insert (TaqI, Sau3a, HinF), none produced the distinctive ladder pattern observed for the PSR repeats.

Sequence analysis of TaqI fragments verified that at least part of the Anv79 clone was composed of a tandem repeat. The nucleotide sequences of four subclones are presented in Fig. 6. The longest insert sequenced, nv79-16, contained the basic 94 by repeat unit next to a similar repeat that had a 16 by deletion. A 9 by deletion was found in the subclone nv79-15. The percentage sequence identity among the four sequenced repeats over the common 69 by sites varied from 80%-90%.

Because NV79 repeats were present on the A chromosomes of *N. vitripennis*, a Southern hybridization was performed to determine whether this family was also present in the two sibling species as well as the more distantly related species *Trichomalopsis dubius*. *T. dubius* is a widespread species in North America that occurs

Fig. 6. Nucleotide sequences of ten cloned non PSR specific repeats. These repeats belong to the shared repeat family NV79 (four sequences) or the A chromosome repeat families NV85 (three sequences) and NV126 (three sequences). Base changes are denoted by letters and deletions are indicated by dashes

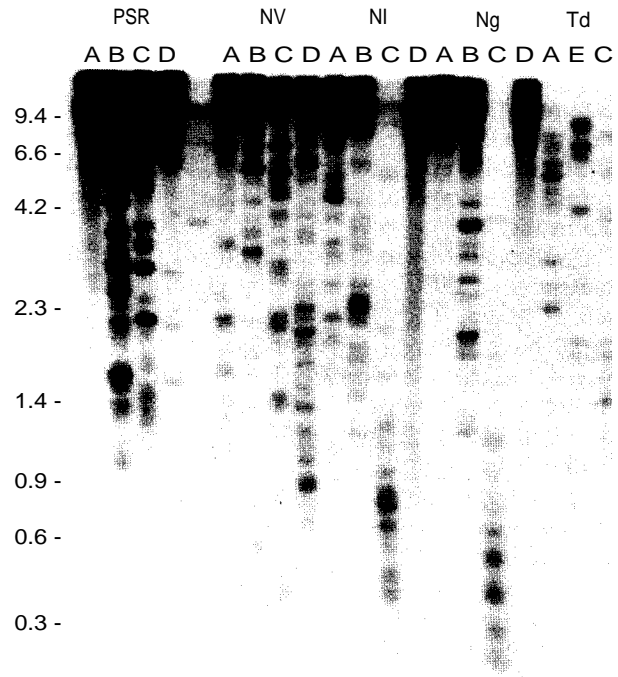


Fig. 7. Genomic blot of DNA from PSR, *Nasonia vitripennis* (Nv), *N. longicornis* (NO, *N. giraulti* (Ng), and *Trichomalopsis dubius* (Td) probed with NV79. The PSR and *N. vitripennis* DNA was isolated from males only while both females and males were used for the remaining species. The PSR lanes contain approximately 0.4  $\mu$ g of DNA while the remaining lanes contain 3  $\mu$ g of DNA digested as follows: A, EcoRI; B, HindIII; C, AluI; D, Aval; E, HaeIII. Hybridization was conducted under moderate stringency conditions. The NV79 probe was the entire insert isolated from nv79. DNA standards are given in kilobases

in the same subfamily (Pteromalini) as *Nasonia* (Boucek 1988). Total genomic DNA from *N. vitripennis* (PSR and wild-type), *N. longicornis*, *N. giraulti* and *T. dubius* was probed with labeled nv79 insert (Fig. 7). Approximately sevenfold less DNA was loaded in the PSR lanes to compensate for the increased level of hybridization observed in preliminary Southern blots. As is apparent from this hybridization experiment, NV79 repeats were found in all four species tested. The level of hybridization in the sibling species was comparable to that in normal *N. vitripennis* but was approximately tenfold less

in *T. dubius*. The absence of the simple ladder pattern observed for the PSR repeats suggested that the degree of sequence diversity was greater in the NV79 family, and the fact that several restriction digests gave rise to large molecular weight DNA suggested that NV79 repeats were arranged in long arrays. It is not known whether they are limited to a single A chromosome.

There are two possible explanations for the higher level of hybridization in PSR males than in wild-type males: (1) NV79 repeats have accumulated on the PSR chromosome relative to the A chromosomes and/or (2) the NV79 repeats have sufficiently diverged that, assuming the clone sequences originated on the PSR chromosome, they hybridized more efficiently to repeats in the PSR lanes. In either case, NV79 repeats were found on both the PSR and A chromosomes, and these repeats were present on the A chromosomes prior to the divergence of *Nasonia* and *Trichomalopsis*.

#### Analysis of A chromosome repeats

Clones were also isolated from the genomic lambda library that showed similar hybridization patterns when probed with labeled PSR or wild-type total genomic DNA, and thus were likely to contain repetitive DNA present on the A chromosomes. Three of these clones, *nv85*, *Anv104*, *Anv126*, were selected for further investigation. These three clones did not cross hybridize to the PSR specific repeats or to *nv79*. *nv85* and *Anv104* weakly cross hybridized with each other at moderate stringency, but neither cross hybridized to *Anv126*. Restriction enzyme analysis indicated that *nv85* and *Anv104* contained a tandem repeat comprising an approximately 170 by monomer repeat unit. In both clones, a ladder was observed consisting of monomer, dimer, and trimer sized fragments when the insert was digested with *Sall*. While not digested to a short repeat, *Anv126* was also assumed to contain tandemly repetitive DNA because ten commonly used restriction enzymes did not cleave the 12 kb insert.

Sequence analysis confirmed that all three clones were composed of short, tandem repeats. Two sequenced repeats from *nv85* were 175 by in length while a repeat from *Anv104* was 162 by in length (Fig. 6). Pairwise comparison revealed 91% sequence identity between *nv85-7* and *nv85-9* while *nv104-6* had approximately 72% sequence identity with both repeats from *nv85*. The two repeat types were considered to be members of different subfamilies of the NV85 repeat family. The three repeats sequenced from *Anv126* were all 110 by in length and had 82%-87% sequence identity (Fig. 6).

Southern analysis of genomic DNA (Fig. 8) comparing equal amounts of DNA from PSR and wild-type males revealed no significant difference between the two sources of DNA, either in intensity of hybridization or banding pattern, for the NV126 and NV85 repeat families. This result indicated that these repeats had not been amplified on the PSR chromosome although the presence of a few repeats on the PSR chromosome was not eliminated. Relative exposure lengths indicated that the

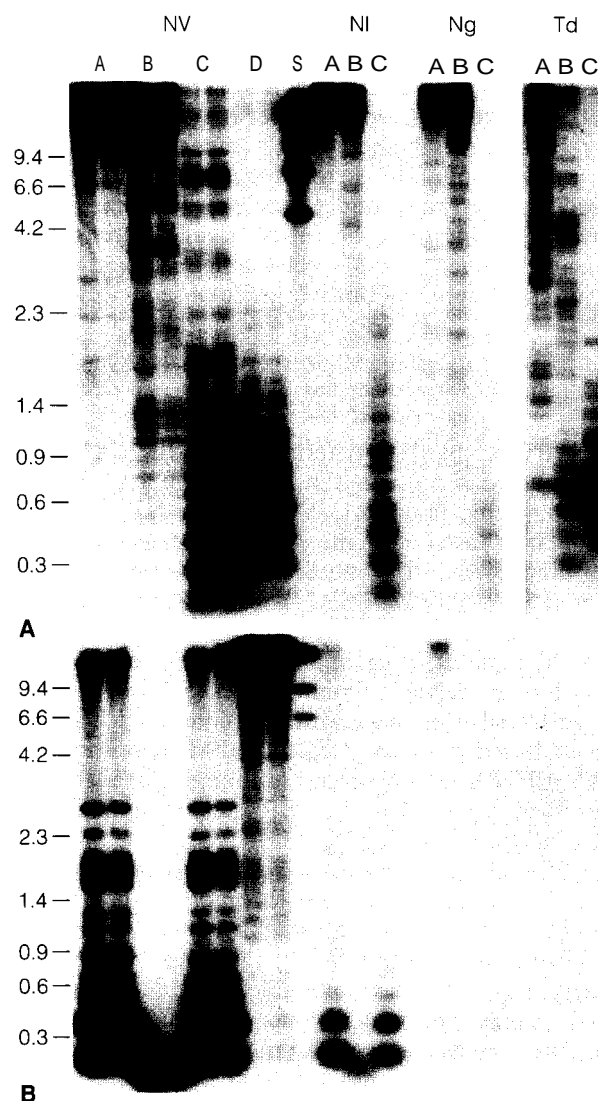


Fig. 8A, B. Blots of genomic DNA from *Nasonia vitripennis* with and without PSR (Nv), *N. longicornis* (NI), *N. giraulti* (Ng), and *Trichomalopsis dubius* (Td) probed with NV126 (A) or NV85 (B). For *N. vitripennis*, the left of each pair contains genomic DNA isolated from males carrying the PSR chromosome while the right lane contains DNA isolated from wild-type males. Hybridizations were done using moderate stringency conditions. Genomic DNAs were digested as follows: for A, A, *EcoRI*, B, *HaeIII*, C, *TagI*, D, *AluI* and for B A, *Sall*, B, *TagI*, C, *HincII*, D, *HaeIII*. In A, exposure for *T. dubius* is ten times longer than for the rest of the Southern blot. DNA standards are given in kilobases at the left

NV 126 repeats were more abundant than the NV85 family. The banding patterns observed in these Southern blots of genomic DNA was more complicated than that seen for the PSR specific families, but for each family one or more enzyme gave rise to a pattern suggestive of a ladder. Presumably, the repeats are arranged in large arrays and this complex pattern reflects restriction site polymorphism among the repeats.

It is apparent from the genomic DNA blots that the NV126 and NV85 repeat families were present in the sibling species and the former was also present in *T. dubius*. NV126 repeats hybridized at approximately 5-

fold lower levels to *N. longicornis* and *N. giraulti* genomic DNA and at approximately 50-fold lower levels to *T. dubius* DNA than to *N. vitripennis* DNA. Hybridization of NV85 repeats, on the other hand, was approximately 10-20 times higher to *N. vitripennis* than to *N. longicornis* genomic DNA, was barely detectable to *N. giraulti* DNA and was not detectable to *T. dubius* DNA. This variation in level of intensity of hybridization among the different wasp species reflects either differences in repeat numbers or divergence in sequence.

## Discussion

A summary of the data determined for the six repeat families studied is presented in Table 1. On average, the non PSR specific families (NV79, NV85, NV126) were found to be shorter and the AT content less variable among families than that observed for the PSR specific families (PSR2, PSR18, PSR22). Matrix analysis comparing the PSR repeat families with the NV126, NV85, and NV79 families revealed no sequence identity over the length of the repeats. Small patches (5-8 bp) of identity were found, but these were not conserved between the A or shared repeat and all three PSR specific families. An AT rich region similar, although not identical, to conserved region II in the PSR specific families was found in all three non PSR families.

The PSR chromosome is of special interest because of its ability to disrupt normal chromosome condensation and/or mitosis, thus resulting in the loss of the paternal chromosome set in the fertilized egg. Two basic alternative mechanisms are possible. First, sequences on the PSR chromosome may produce a product (e.g. DNA binding protein, methylase, etc.) that imprints the paternal chromosomes in such a way that they dysfunction. Second, sequences on PSR may act as a "sink" that bind a product(s) necessary for normal chromosome processing. Although supercondensation of the paternal chromosomes occurs at the first mitotic division of the fertilized egg, modification of the paternal chromosomes may actually have occurred in the male during spermatogenesis. This seems likely because there is a short time period (approximately 30 min) between fertilization and compaction of the paternal set, which would not allow

much time for PSR to function upon entering with the sperm.

Three families of short, tandemly repetitive DNA (PSR2, PSR18, PSR22) were localized to the PSR chromosome. At least one-third of the PSR chromosome is estimated to comprise these repeat families (Nur et al. 1988). The PSR2 family appeared to be the most highly conserved as well as the most abundant of the three, which may be the result of a recent amplification. The three families shared two conserved regions and were likely to have arisen from a common ancestor. Also present on the PSR chromosome was the shared repeat family NV79. The relative position of these four families on the PSR chromosome is not known. Southern and genetic (L. Beukeboom, in preparation) analysis indicated that each family was present on the PSR chromosome in a single array. Four clones have been isolated that hybridize to both the PSR2 and PSR18 families, and a single clone has been found that hybridizes to both the PSR18 and PSR79 families. This suggests that the PSR18 family is bordered by the PSR2 and PSR79 families. Attempts are now in progress to determine the nature of the sequences at the junctions between these families.

What makes the PSR repeat families interesting is that their presence is limited to essentially a single chromosome. The mode of PSR inheritance is at least partially responsible for this localization. Any sequences transferred from the PSR to the A chromosomes by non-homologous recombination or translocation will not be transmitted because the A chromosomes are destroyed by PSR at every generation. This allows the repeats on the PSR chromosome to evolve semi-independently of the autosomes. On the other hand, acquisition of sequences from the autosomes on the PSR could occur.

It is not known whether the PSR specific repeat families have a function in either the abnormal condensation of the A chromosomes or in protecting the PSR chromosome from its own action. The satellite sequences may be functionally inert, i.e. present solely as part of the superstructure of the chromosome. However, it is difficult to explain how the palindromic sequences have remained conserved in comparison with the flanking regions despite there being tens of thousands of copies within the genome unless they have some functional significance. One possibility is that the conserved region I functions as a binding site for a protein necessary in paternal A chromosome condensation and/or mitotic division. Palindromes have frequently been determined to play an important role in protein binding (Halazonetis and Kandil 1991; Lee et al. 1987). A second possibility that the data suggest is unlikely is that the repeats code for a small protein that functions in one of these processes. Both the PSR2 and PSR18 families have open reading frames that could code for 36 (PSR2) or 46-51 (PSR18) amino acid proteins. The presence of a TATAA like segment found in common gene sequences approximately 35 by upstream (conserved region II) is interesting, however, the coding frame is not maintained between the two repeat families. In addition, RNA blotting analysis did not reveal transcripts homologous to the

Table 1. Summary of data obtained for the six repeat families

Repeat family	Length (bp)	%AT	PSR <sup>a</sup>	Nv	Nl	Ng	Td
PSR2	171	54	+	-	-	-	nd
PSR18	152-214	61	+	-	-	-	nd
PSR22	183	66	+	-	-	-	nd
NV79	78-94	59	+	+	+	+	+
NV85	162-175	61	-	+	+	+	-
NV126	110	56	-	+	+	+	+

nd, not done; Nv, *Nasonia vitripennis*; Nl, *N. longicornis*; Ng, *N. giraulti*; Td, *Trichomalopsis dubius*

<sup>a</sup> Presence of the repeat on the B chromosome alone



PSR2 or PSR18 repeats (data not shown). TATAA like sequences have also been found in the complex satellite DNAs of *Xenopus* (Lam and Carroll 1983) and *Drosophila* (Miklos and Gill 1981); however, no biological significance has been attributed to them. Alternatively, the palindromic sequences may not be involved in PSR induced condensation of the A chromosomes but rather could facilitate unequal pairing and therefore amplification of the tandem arrays as has been proposed for palindromes found in repeats of two other wasp species (Bigot et al. 1990).

Although the origin of the PSR chromosome is unknown, it is a reasonable assumption that most B chromosomes derived from one of the normal A chromosomes (Jones and Rees 1982). Therefore, it should not be surprising that the PSR chromosome was found to have sequences in common with the A chromosomes. The shared repeat NV79 was interesting for two reasons. First it was present at relatively high levels in the two sibling species *N. longicornis* and *N. giraulti* as well as the more distantly related species *T. dubius*, which suggested it has been around since the divergence of *Trichomalopsis* from the *Nasonia* genus. Unfortunately, there are currently no data on the relative times of divergence of these species. Secondly, it was apparently greatly amplified on the PSR chromosome. Comparison of NV79 sequences derived from the A chromosomes from *N. vitripennis* as well as from *N. giraulti* and *N. longicornis* may provide a clue as to the origin of the PSR chromosome.

Finally, these experiments revealed that the satellite DNA in all three *Nasonia* sibling species as well as *T. dubius* was composed of several different families. Three different families of satellite DNA were identified in wild-type *N. vitripennis*. Additional clones containing repetitive DNA have been isolated that do not cross hybridize with any of the characterized families, which suggests that additional families may exist (data not shown). In addition, two out of the three species of satellite DNA found in *N. vitripennis* were also detected in *Trichomalopsis*. This was in marked contrast to the results found for two other species of parasitic wasps (Bigot et al. 1990). Satellite DNA in *Diadromus pulchellus* and *Eupelmus vuilleti* was determined to consist of a single family of DNA unique to that species.

Presently experiments are underway to isolate additional sequences, both repetitive and unique, present on the PSR chromosome. Genetic experiments are in progress to determine the region of the PSR chromosome necessary for function. By combining genetic and molecular genetic approaches, we hope eventually to determine the sequences responsible for PSR action and to determine the origin of the PSR chromosome.

*Acknowledgements.* We thank Bill Burke for assistance with the DNA sequencing and Leo Beukeboom for discussions and for the

use of unpublished data. This investigation was supported by the NIH.

## References

- Benton WD, Davis RW (1977) Screening .lgt recombinant clones by hybridization to single plaques in situ. *Science* 196:180-182
- Bigot Y, Hamelin M-H, Periquet G (1990) Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). *Mol Biol Evol* 7:351-364
- Boucek Z (1988) Australian chalcidoidea [Hymenoptera]. CAB International, Wallingford, UK
- Breeuwer JAJ, Werren JH (1990) Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. *Nature* 346:558-560
- Darling DC, Werren JH (1990) Biosystematics of *Nasonia* (Hymenoptera:Pteromalidae): Two new species reared from birds' nests in North America. *Ann Entomol Soc Am* 83:352-370
- Halazonetis TD, Kandil AN (1991) Determination of the c-MYC DNA-binding site. *Proc Natl Acad Sci USA* 88:6162-6266
- Jones RN, Rees H (1982) B Chromosomes. Academic Press, New York
- Kreitman M, Hudson R (1991) Inferring the evolutionary histories of the *Adh* and *Adh-dup loci* in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* 127:565-582
- Lam BS, Carroll D (1983) Tandemly repeated DNA sequences from *Xenopus laevis* I. Studies on sequence organization and variation in satellite 1 DNA (741 base-pair repeat). *J Mol Biol* 165:567-585
- Lee W, Haslinger A, Karin M, Tjian R (1987) Activation of transcription by two factors that bind promoter and enhancer sequences of human metallothionein gene and SV40. *Nature* 325:368-372
- Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Miklos GLG, Gill AC (1981) The DNA sequences of cloned complex satellite DNA From Hawaiian *Drosophila* and their bearing on satellite DNA sequence conservation. *Chromosoma* 82:409-427
- Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH (1988) A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome. *Science* 240:512-514
- Ryan SL, Saul II GB, Conner GW (1985) Aberrant segregation of R-locus genes in male progeny from incompatible crosses in *Mormoniella*. *J Hered* 76:21-26
- Ryan SL, Saul II GB, Conner GW (1987) Separation of factors containing R locus genes in *Mormoniella* stocks derived from aberrant segregation following incompatible crosses. *J Hered* 78:273-275
- Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467
- Werren JH (1991) The PSR (paternal sex ratio) chromosome. *Am Nat* 137:392-402
- Werren JH, van den Assem J (1986) Experimental analysis of a paternally inherited extrachromosomal factor. *Genetics* 114:217-233
- Werren JH, Skinner SW, Charnov EL (1981) Paternal inheritance of a daughterless sex ratio factor. *Nature* 293:467-468
- Werren JH, Nur U, Eickbush D (1987) An extrachromosomal factor causing loss of paternal chromosomes. *Nature* 327:75-76