# Physical Analysis of the Campoletis sonorensis Virus Multipartite Genome and Identification of a Family of Tandemly Repeated Elements

DAVID A. THEILMANN AND MAX D. SUMMERS\*

Department of Entomology, Texas A&M University, and Texas Agricultural Experiment Station, College Station, Texas 77843-2475

Received 3 February 1987/Accepted 2 May 1987

This report is an analysis of cross-hybridizing sequences found within the 28 superhelical (SH) DNAs of the multipartite genome of the polydnavirus *Campoletis sonorensis* virus (CsV). A Southern cross-blot hybridization analysis showed that the majority of CsV *Eco*RI restriction fragments cross-hybridize to multiple *Eco*RI fragments. These sequence homologies were analyzed by hybridizing recombinant clones of the CsV SH DNAs B, H, M, and O<sup>1</sup> to Southern blots of undigested CsV DNA, using different hybridization stringencies. The results indicated that homologous regions among the SH DNAs include closely related sequences that are detectable under stringencies. A sequence that hybridized to the majority of the CsV SH DNAs was identified and subcloned from the SH DNAs O<sup>1</sup>, H, and B. Nucleotide sequence data revealed that these homologous regions contained a family of imperfectly conserved repeated elements. These repeat elements were arranged singly or in direct tandem arrays and had an average length of 540 base pairs. Within the sequenced regions that contained the repeated elements six putative open reading frames were identified. These results show that the CsV genome consists of SH DNAs with complex sequence interrelationships that may have arisen due to multiple recombinational events.

Campoletis sonorensis virus (CsV; Polydnaviridae) replicates in the calyx epithelium of the reproductive tissue of the female parasitic wasp C. sonorensis (Ichneumonidae) (26). CsV is secreted into the oviduct lumen and during oviposition is coinjected with the wasp egg into the hemocoel of the lepidopteran host Heliothis virescens (Noctuidae) (26). If the wasp egg is artificially injected into H. virescens without CsV, the egg will be encapsulated by the host hemocytes (11). In the presence of CsV, the parasite egg will not be encapsulated and the egg will hatch, with the eventual result being the death of the host and the development of a C. sonorensis adult (11, 26). Thus it appears that by some presently unknown mechanism, CsV protects the parasite egg and is therefore required for the successful parasitism of H. virescens by C. sonorensis.

The injection of calyx fluid or purified CsV into H. virescens, which mimics natural parasitization, results in dramatic effects upon the development and physiology of the host. These effects include reduced growth rate and retarded development (27–29), changes in hormone titers (9a), and alterations to the number and behavior of plasmatocytes (7a). Molecular analyses of CsV transcription in H. virescens show that CsV transcripts are detected during the entire development of the endoparasitic wasp larva and that at least 12 different mRNAs can be identified (2, 3, 12). During this same period, CsV persists in H. virescens but no viral replication is detected (27).

The genome of CsV consists of at least 28 superhelical (SH) DNAs that range in size from 6 to greater than 20 kilobase pairs (kbp) (2, 16). In the initial analysis of the CsV genome, sequence relationships among the individual SH DNAs were analyzed by hybridizing labeled probes of gel-purified SH DNAs to Southern blots of CsV DNA (16).

Because only limited cross-hybridization between the SH DNAs was detected, it was concluded that the CsV genome comprised mostly unique sequences. However, more recent data showed that when recombinant clones of CsV DNA were hybridized to the CsV genome, homology to multiple SH DNAs was observed (3, 13, 27). For example, Blissard et al. (3) showed that a CsV clone (p2H-5300), which contains sequences homologous to a viral mRNA, hybridizes strongly to at least 11 different SH DNAs. These results indicate that some of the SH DNAs of the CsV genome do not contain mostly unique sequences and that some DNA sequences were repeated on more than one SH DNA. In the present investigation, we conducted a more detailed study of the physical organization of the viral genome specifically to determine if the repetition of sequences was more common throughout the CsV genome than previously thought. In addition we analyzed sequence relationships between specific cloned SH DNAs and the rest of the CsV genome.

Our Southern cross-blot hybridization analysis demonstrated that the majority of the CsV EcoRI fragments hybridized to at least one region of the CsV genome other than themselves and usually to more than one. Cross-hybridization of individual CsV SH DNAs was analyzed by hybridization of cloned SH DNAs to the CsV genome under various hybridization stringencies. The results showed that each of the cloned SH DNAs analyzed contain sequences that cross-hybridize to other SH DNAs. Some of the crosshybridizing regions could only be detected under conditions of reduced stringency. Fragments that cross-hybridize to most of the SH DNAs in the CsV genome were isolated and cloned from three separate SH DNAs. The nucleotide sequences of these homologous regions were determined and shown to consist of a family of repeated elements that were arranged singly or in tandem arrays. The results of this study indicated that the CsV genome comprises interrelated SH

<sup>\*</sup> Corresponding author.

DNAs that may have arisen by multiple recombinational events and sequence divergence.

#### MATERIALS AND METHODS

Insect cultures and virus purification. The insects H. virescens and C. sonorensis were reared at 27°C on a 14-h-10-h (light:dark) photoperiod as described by Krell et al. (16). Third-instar H. virescens were parasitized by mated female C. sonorensis wasps as described by Fleming et al. (12).

CsV was purified from dissected female wasp oviducts by the method of Krell et al. (16). Viral DNA was extracted from purified virions by using the procedure described by Fleming et al. (12).

Southern blots. DNA samples were separated by agarose gel electrophoresis and transferred to nitrocellulose membranes by the procedures of Southern (25) and Smith and Summers (24).

[<sup>32</sup>P]dATP-labeled probes (22) were hybridized to Southern blots by using various hybridization stringencies by the procedures of Howley et al. (15). Hybridizations of labeled probes were performed at 43°C in 50, 40, or 30% formamide (untreated; Bethesda Research Laboratories, Inc., Gaithersburg, Md.)-150 mg of calf thymus DNA per ml-5× Denhardt solution (0.02% Ficoll [Sigma Chemical Co., St. Louis, Mo.], 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin)-20 mM sodium phosphate (pH 6.5)-5× SSC (0.75 M NaCl, 0.075 M sodium citrate). The blots were washed three times in  $2 \times$  SSC-0.1% sodium dodecyl sulfate at room temperature for 10 min, followed by three 30-min washes in the low salt buffer  $(0.1 \times SSC, 0.1\%)$  sodium dodecyl sulfate) at temperatures equivalent to the stringencies of the hybridization conditions. These temperatures were determined by using the equation  $T_m = 81.5 + 16.6 (\log [Na^+]) + 0.41 (\%$ G+C)-0.72 (% formamide) (15). A G+C content of 50% was previously determined by Krell et al. (16) and was used in the calculations. Blots hybridized in 50, 40, or 30% formamide were washed at 50, 43, or 37°C, respectively. Under these conditions the 50, 40, and 30% formamide hybridization buffers will permit hybridization when homology is greater than 77, 70, and 64%, respectively, assuming that a 1% mismatch lowers the  $T_m$  by 1.0°C (19).

Southern cross-blot hybridization analysis. The Southern blot cross-hybridization was performed with GeneScreen and GeneScreen-Plus membranes by the procedure of the manufacturer (New England Nuclear Corp., Boston, Mass.). Briefly, CsV DNA was completely digested with the restriction enzyme EcoRI. The digest was separated into two portions, and one portion was 3' end labeled with [32P]dATP by using the Klenow fragment of DNA polymerase I. The labeled (2  $\times$  10<sup>4</sup> dpm) and unlabeled (1 µg) EcoRI CsV digests were separated in parallel in two 0.7% agarose gels at 55 V for 15 h by using single, wide preparatory wells. Control lanes of the labeled and unlabeled digests were electrophoresed alongside the wide sample lanes. The separated unlabeled and labeled digests were transferred to GeneScreen-Plus and GeneScreen membranes, respectively, by capillary transfer as described by the manufacturer. The DNA was fixed to the GeneScreen-Plus membrane, placed at right angles to the labeled blot, immersed in hybridization buffer (0.5 M NaCl, 10% dextran sulfate, 50% formamide, 150 mg of calf thymus DNA per ml), sandwiched between two glass plates, and hybridized overnight at 43°C. The GeneScreen-Plus membrane was washed in  $2 \times$  SSC-0.1% SDS at room temperature followed by washes in  $0.1 \times$ 

SSC-0.1% sodium dodecyl sulfate at 50°C. The blot was then air dried and exposed to X-ray film (XAR-5; Eastman Kodak Co., Rochester, N.Y.) with an intensifying screen (Cronex Lightning-Plus; E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.) at -80°C.

**Sequencing.** The sequencing reactions were performed by using the dideoxy-nucleotide procedure of Sanger et al. (23). DNA fragments to be sequenced were cloned into the bacteriophage vectors M13 mp18 and mp19 (31). Overlapping sets of deletions were generated by using the double-stranded DNA *ExoIII-ExoVII* procedure described by Yanisch-Perron et al. (31) and the single-stranded DNA T4-DNA polymerase technique of Dale et al. (7). The sequences of both strands of all DNA fragments were determined.

Sequence data analysis was performed with the University of Wisconsin Genetics Computer Group DNA analysis software for the VAX minicomputer (8, 18).

Cloning of CsV DNAs. Standard methods used for the isolation, cloning, and purification of DNAs are described in Maniatis et al. (19). Clones of specific CsV SH DNAs were obtained by band isolating individual SH DNAs from 0.7% agarose gels in which undigested viral genomic DNA had been electrophoretically separated. The SH DNAs are named according to the previously established nomenclature (2, 16). The isolated DNAs were digested with different restriction enzymes, and the fragments were cloned into the plasmid vectors pUC8 or pUC18. Overlapping restriction maps of the cloned fragments were used to identify complete SH DNAs. CsV SH-B had been previously cloned (27). Single insert clones of SH-O<sup>1</sup> and -M were isolated from an existing pBR325 SalI shotgun library of the CsV genome after the SH-DNAs were mapped from band-isolated subclones. The plasmid clones of the CsV SH DNAs B, M, and O<sup>1</sup> will be referred to as pSH-B, pSH-M, and pSH-O<sup>1</sup> respectively. The 8.4-kbp HindIII clone of CsV SH-H was previously identified as p4H-8460 (2, 3). This CsV SH-H clone is believed to contain the complete sequence of the CsV SH-H DNA and will be referred to as pSH-H.

## RESULTS

To examine the overall cross-hybridization throughout the CsV genome a Southern cross-blot hybridization was performed by using an EcoRI digest of CsV genomic DNA. This procedure is similar to individually hybridizing each CsV EcoRI restriction fragment to a Southern blot of an EcoRI digest of the entire CsV genome. An EcoRI digest was used because the resulting fragments were distributed over a wide range of sizes, allowing the resolution of 54 bands (Fig. 1). Analysis of the Southern cross-blot hybridization (Fig. 1) shows an obvious diagonal of dark spots from the top left-hand corner to the bottom right-hand corner which represents the hybridization of each restriction fragment to itself. Spots off the diagonal represent the cross-hybridization between restriction fragments of different sizes. The regions above and below the diagonal are mirror images; variation in spot intensity on either side is due in part to differences in the sizes and molar amounts of the end-labeled probe fragments. The large number of spots occurring off the diagonal show that numerous cross-hybridizing sequences are found throughout the CsV genome. A majority of the 54 resolvable bands (48) cross-hybridized to at least one different band and usually to more than one (Fig. 1). We expected the large fragments to cross-hybridize to more of the DNA fragments because large fragments could poten-



FIG. 1. Southern cross-blot hybridization of EcoRI-digested CsV DNA. The top panel is the ethidium bromide-stained marker lane of the unlabeled CsV digest. The numbers refer to the individual EcoRI fragments, with 1 being the largest and 54 the smallest detectable band. The side panel shows an autoradiogram of the [<sup>32</sup>P]dATP end-labeled EcoRI digest of CsV DNA and of end-labeled molecular weight markers (*Hind*III digest of lambda phage DNA). The arrowheads at the top and side panels refer to band 34. The details of this procedure are described in the text.

tially contain a greater number of cross-hybridizing sequences. However, smaller DNAs also cross-hybridized to many restriction fragments. For example, band 34 (Fig. 1, arrows) is less than 2 kbp but hybridizes to bands 49, 46, 41, 32, 31, 30, 28, 24, 25, 22, 21, 19, 15, 11, 9, 8, 7, 4, and 1.

The Southern blot cross-hybridization analysis is a qualitative assessment of cross-hybridizing sequences throughout the CsV genome. Since a restriction digest of the CsV genome is used in this procedure and the origin of most of the CsV *Eco*RI fragments is unknown, the results do not specifically illustrate the nature of these sequence relationships. To analyze the sequence relationships between individual SH DNAs, recombinant clones of four CsV SH DNAs of known origin were used for analysis of cross-hybridizing sequences. They were pSH-B (6.6 kbp), pSH-H (8.4 kbp), pSH-M (10.8 kbp), and pSH-O<sup>1</sup> (11.2 kbp). These DNAs were chosen because they represent a wide range of the SH DNA sizes present in the CsV genome (Fig. 2A).

Figure 2B shows the hybridization of cloned SH DNAs to Southern blots of CsV genomic DNA under various conditions of hybridization stringency. Each of the cloned SH DNAs hybridized to multiple SH DNAs other than itself. The number of SH DNAs to which clones pSH-B, -H, and  $-O^1$  cross-hybridized increased as the stringency of hybridization was decreased. SH-B cross-hybridized intensely to SH-Q and  $-A^1$  under the lower stringencies of 40 and 30% formamide; minor cross-hybridization to five other SH DNAs was also observed (Fig. 2B). Under stringent conditions, pSH-H hybridized mainly to itself and showed only minor hybridization to SH-L<sup>1</sup> and -Q. However, as the stringency was reduced hybridization to SH-O<sup>1</sup>, -R, and -T was also easily detected. A similar hybridization pattern has been previously reported for pSH-H (3). Unlike the other three SH DNAs tested, hybridization of pSH-M under conditions of reduced stringencies did not identify additional SH DNAs that contained homologous sequences.

Of the four cloned SH DNAs analyzed,  $pSH-O^1$  crosshybridized to the most DNAs. Under stringent conditions SH-O<sup>1</sup> hybridizes with an almost equal or greater intensity to SH-Q, -R and -U as it does to itself (Fig. 2B). These results indicated that the four SH DNAs O<sup>1</sup>, Q, R, and U contain some sequences that are very closely related. Under lowstringency conditions (40 and 30% formamide), it was observed that SH-O<sup>1</sup> also hybridized at lower intensities to nearly every SH DNA. This includes DNAs that migrate below SH-A<sup>1</sup>, at which no DNAs are visually detectable on the ethidium bromide-stained control lane (Fig. 2A). Low levels of hybridization to the majority of the SH DNAs could also be observed if the blot hybridized in 50% formamide was exposed to X-ray film for long periods.

Identification of repeated sequences. The low-level hybridization of SH-O<sup>1</sup> to a majority of the SH DNAs in the CsV



FIG. 2. Hybridization analysis of cloned CsV SH DNAs. (A) Ethidium bromide-stained 0.7% agarose gel of electrophoretically separated undigested CsV genomic DNA (1  $\mu$ g). The locations of the cloned CsV SH DNAs are indicated on the right. The lines on the left indicate the regions of the relaxed circular (RC) and SH forms of the CsV DNAs. (B) Southern blots of duplicate lanes from the same gel shown in panel A were hybridized with pSH-B, -H, -M, and -O<sup>1</sup> as indicated. The stringency of the hybridization used, expressed in percent formamide, is shown above each lane. The probe concentration and specific activity and autoradiograph exposures times were the same for the three stringencies used for each of the individual cloned CsV SH DNAs. The SH DNAs to which each of the cloned probes hybridized are shown on the right side of each section. The RC forms of each probe are shown. Markers with no letter designation indicate bands for which there is no corresponding previously identified band in the ethidium bromide-stained control lane.

genome indicated that this molecule contained a sequence or sequences that were repeated on most of the SH DNAs. To map the locations of these repeated sequences on pSH-O<sup>1</sup>, 17 overlapping restriction fragments of this molecule were gel purified, labeled, and hybridized to Southern blots of CsV genomic DNA (Fig. 3). Figure 3A shows the location and size of each probe used beneath a map of pSH-O<sup>1</sup>. On the basis of the intensity of hybridization, one of two hybridization profiles was observed for all the probes tested. Of the 17 probes, 12 hybridized intensely to SH-O<sup>1</sup>, -Q, -R, and -U and less intensely to the majority of the other SH DNAs and therefore were scored as containing the repeat (Fig. 3B, Repeat Positive). This was the same hybridization profile as that seen for the complete clone pSH-O<sup>1</sup> (Fig. 2). The other hybridization profile observed is also shown in Fig. 3B. Of the 17 probes, 5 hybridized intensely to SH-O<sup>1</sup>, -Q, -R, and -U but did not have the less intense hybridization to the majority of the SH DNAs in the CsV genome. The fragments that gave this hybridization profile were scored as not containing the repeated sequences (Fig. 3B, Repeat Negative). The probes containing the repeated sequences mapped to two locations on SH-O<sup>1</sup> that were 1.1 kbp (Fig. 3A, region II) and 5.6 kbp (regions I and III) in size. The cumulative size of the repeat-positive fragments is 6.7 kbp, which accounts for approximately 60% of pSH-O<sup>1</sup>.

To identify the regions on other SH molecules which hybridize to the repeated sequences of SH-O<sup>1</sup>, the smaller 1.1-kbp region of SH-O<sup>1</sup> (Fig. 3A, region II) was subcloned into pUC8 and this clone, pO<sup>1</sup>-HC1185, was used as a probe to map homologous sequences on the cloned SH DNAs B, H, and M. Clone pO<sup>1</sup>-HC1185 was hybridized to Southern blots of restriction enzyme digests of clones pSH-O<sup>1</sup>, pSH-H, pSH-B, and pSH-M. The regions on these molecules to which pO<sup>1</sup>-HC1185 hybridized are shown in Fig. 4. On pSH-H a region homologous to pO<sup>1</sup>-HC1185 was mapped to a 2.7-kbp region. On pSH-B, pO<sup>1</sup>-HC1185 hybridized to a 1.0-kbp region but at very low levels. No hybridization of pO<sup>1</sup>-HC1185 to pSH-M was detected. When pO<sup>1</sup>-HC1185 was hybridized to restriction digests of the parent molecule pSH-O<sup>1</sup> hybridization was not detected to repeat region III, even under reduced (40% formamide) stringencies. This suggested that at least two types or classes of repeated sequences may be present on SH-O<sup>1</sup> because region III had been shown to contain repeated sequences by our hybridization assay (Fig. 3).

Nucleotide sequence analysis of repeat regions. Comparisons of the homologous regions identified on SH-B, -H, and  $-O^1$  did not reveal any common restriction sites (Fig. 4). Therefore, the nucleotide sequence of  $pO^1$ -HC1185 and the homologous regions on pSH-B, and pSH-H were determined. The regions sequenced and the strategy used are shown in Fig. 4.

The regions sequenced on pSH-O<sup>1</sup>, -H, and -B were 1,303, 2,743, and 1,153 bp in length, respectively. Dot matrix homology comparisons were used initially to analyze the three regions to identify homologous sequences (Fig. 5). The diagonal lines of a dot-plot comparison represent regions of homology, spaces in the lines indicate regions of sequence mismatch relative to the window size and stringencies used. The five major broken diagonal lines of the dot-plot comparison in Fig. 5A indicate that sequences on pSH-B between nucleotides (nt) 44 and 587 (BR1) are imperfectly repeated approximately 4.5 times (HR1 through HR5) in a direct tandem arrangement on the pSH-H molecule. A similar comparison between the pSH-O<sup>1</sup> and pSH-B sequences (Fig. 5C) indicates that the same region of pSH-B that is repeated on pSH-H is also repeated on pSH-O<sup>1</sup> approximately 2.5 times (OR1 through OR3) and that the sequences are also arranged in a direct tandem array. Comparison of the repeated sequences of pSH-H and pSH-O<sup>1</sup> (Fig. 5B) results in a complex pattern of five long overlapping diagonal lines that are due to the homology between the 4.5 repeats of pSH-H and the 2.5 repeats of pSH-O<sup>1</sup> which are organized in tandem arrays.

The dot plots showed that the sequences from the three SH DNAs share homology due to a repeated element present in 1, 4.5, and 2.5 copies on the sequenced regions of pSH-B, pSH-H, and pSH-O<sup>1</sup>, respectively. The broken diagonal lines show that the homology between the individual repeated elements is not perfect. To analyze the relative



FIG. 3. Mapping of repeat sequences on the cloned SH-O<sup>1</sup>. (A) The locations of the 17 gel-purified restriction fragments used as hybridization probes to map the repeat sequences are shown beneath the restriction map of pSH-O<sup>1</sup>. The consensus locations of the repeat sequences determined by the probes are indicated by the solid black lanes on the restriction map (regions I, II, and III). Regions I and III would be joined in the circular molecule. (B) Hybridization patterns of the probes shown in panel A that either contained the repeated sequences (REPEAT POSITIVE) or did not contain the repeated sequences (REPEAT NEGATIVE). The labeled gel-purified probes were hybridized to Southern blots of undigested CsV genomic DNA (1  $\mu$ g) under stringent conditions (50% formamide). The minor hybridization to multiple SH DNAs that was specific to the repeat-positive probes is indicated by the arrow. For both the repeat positive and repeat negative panels, lanes 1 and 2 represent 12- and 30-h autoradiograph exposures, respectively.

sequence composition of all the partial or complete repeat sequences from pSH-B, -H, and  $-O^1$ , all were optimally aligned by computer and a consensus sequence was determined (Fig. 6).

The average length of each repeat element is approximately 540 bp, with each repeat approximately 70% homologous to the consensus repeat. The homology between individual repeats varies from 60 to 70%, with short segments (less than 50 bp) having 80 to 100% homology. The repeat element HR5 contains a sequence of 78 bp that does not share homology with any of the other repeats (Fig. 6). The reduced or partial repeat elements on the ends of the sequenced regions of pSH-O<sup>1</sup> (OR1 and OR3) and pSH-H (HR1) suggest that these repeat sequences extend beyond the regions analyzed. Analysis of the aligned sequences indicated that there are no specific regions within the repeat element that appear to be highly conserved among all the repeats.

Analysis of the sequence data from the three homologous regions on SH-H,  $-O^1$ , and -B has also revealed the presence of six open reading frames (ORFs) which are shown schematically in Fig. 7. The ORFs range in size from 291 to 996 nt. On pSH-H, three ORFs have been identified, SH-O<sup>1</sup>

contains one ORF that starts at nt 326 and continues through to the end of the sequenced region (853 nt), and the pSH-B repeat sequence contains two small ORFs. Comparison of the computer-generated translation products of these ORFs shows homologies among the predicted proteins. Alignment of the predicted protein sequences of H ORF 1 and H ORF 2 identified a region of 70 amino acids that was 69% homologous (data not shown). Similar results were obtained when each of the other predicted proteins from the six ORFs were compared with each other.

## DISCUSSION

The polydnavirus family is the only group of doublestranded DNA viruses that has segmented genomes (4). The CsV genome has at least 28 circular SH DNAs that range from approximately 6.0 to greater than 20 kbp, with a composite size greater than 270 kbp (2, 16). The large genome size and its segmented organization makes CsV one of the largest and most structurally complex of the known animal viruses. Preliminary analysis of the CsV genome suggested that all of the SH DNAs were for the most part unique in sequence content (16). Contrary to this, recent



FIG. 4. Mapping and sequencing strategies of the regions homologous to the repeat plasmid pO<sup>1</sup>-HC1185. Restriction maps of the four cloned SH DNAs (pSH-O<sup>1</sup>, -H, -B, and -M) are shown, with the regions homologous to pO<sup>1</sup>-HC1185 (determined under stringent conditions [50% formamide]) indicated by solid black lines. Sequenced regions of pSH-B, -H, and -O<sup>1</sup> and the sequencing strategies used, as indicated by the arrows, are shown below each map. The regions on SH-O<sup>1</sup> shown to contain repeated sequences (Fig. 3) are indicated (I, II, and III).

data have shown that cloned fragments of CsV DNA will cross-hybridize to as many as 11 different SH DNAs (3, 13, 27). In the present study we performed a more detailed analysis of the physical relationships among the multiple SH DNAs that constitute the CsV genome.

To assess the degree of cross-hybridizing sequences throughout the CsV genome we performed a Southern cross-blot hybridization (Fig. 1) which indicated that sequence homology exists between the majority of the SH DNAs and that most of the SH DNAs are not composed of unique sequences as previously suggested. This result reveals that the CsV genome is composed of interrelated SH DNAs.

Hybridization of DNAs under different stringencies has been used to analyze various degrees of sequence homologies and evolutionary relatedness among viral genomes, gene families, or specific DNA sequences (1, 6, 14, 15). In this study we used variation in hybridization conditions to analyze the sequence relationships among several SH DNAs of the CsV genome. Our hybridization data with four cloned SH DNAs (Fig. 2) revealed that relationships among different DNAs of the CsV genome include both closely related DNA sequences for which homology is detected under stringent conditions and related but diverged sequences which are more easily detected under reduced stringencies.

Hybridization of 17 restriction fragments of SH-O<sup>1</sup> indicated that every region tested hybridized intensely to the molecules O<sup>1</sup>, Q, R, and U under stringent conditions (Fig. 3), which suggests that the SH- $O^1$  molecule is duplicated within these larger SH DNAs. It is possible that the larger SH DNAs (Q, R, and U) have undergone intramolecular recombination to generate smaller circular molecules that will be duplications of portions of the larger molecules. Palmer and Shields (20) reported that the 218-kbp plant mitochondrial genome of Brassica campestris appears to recombine via intramolecular recombination between direct repeats to generate smaller circular DNAs which are duplications of part of the parent mitochondrial molecule (20). This type of recombination may explain the cross-hybridization results we observed among  $SH-O^1$ , -Q, -R, and -U. Cloning and mapping of the larger CsV SH DNAs Q, R, and U will be required to determine whether they do indeed contain all the sequences of SH-O<sup>1</sup>. A 7.2-kbp CsV clone from the CsV-Q region has already been cloned and analyzed (2, 3, 13), but the restriction map of this clone shows no similarity to that of SH-O<sup>1</sup>, which suggests that there may be comigrating bands in the Q region of the CsV genome.



FIG. 5. Dot matrix homology plot comparisons of the repeat sequences from (A) pSH-H and pSH-B, (B) pSH-H and pSH-O<sup>1</sup>, and (C) pSH-O<sup>1</sup> and pSH-B. The stringency of the comparison was a minimum 14-bp match in a window of 21 bp. The numbers on the axes refer to nucleotide base pairs. The locations of the repeat elements pSH-H repeat 1 (HR1) to pSH-H repeat 5 (HR5), pSH-O<sup>1</sup> repeat 1 (OR1) to pSH-O<sup>1</sup> repeat 3 (OR3), and pSH-B repeat 1 (BR1) are shown on the axes.

In this study we also identified, cloned, and sequenced a region of  $pSH-O^1$  that cross-hybridizes under low stringency to the majority of the SH DNAs in the CsV genome. The sequence data from  $pSH-O^1$  and homologous sequences from pSH-B and -H revealed that these three regions consisted of an imperfectly conserved repeated element with an average size of 540 bp. On  $pSH-O^1$  and pSH-H the repeat elements were organized in tandem arrays, but only a single element was identified on pSH-B (Fig. 4). Comparisons of all

the sequenced repeat elements suggest that no specific sequences of the repeat elements are highly conserved (Fig. 6). The sequence data showed that the homology among the repeated elements was variable, with an average homology of 60 to 70%, but that the repeats also contained short regions that were greater than 90% homologous. This explains in part why we observed increased cross-hybridization when low-stringency conditions were used (Fig. 2,  $pSH-O^1$ ). Analysis of the consensus sequence that was

1001	-	CARCERCEINC	MAXAMENT C	AATACTTAAA	*****	ממדמידידיאמ	111111111111111111111111111111111111111	ATAGAAGETG	GAATACAAGT	TOGACAGOGA
HR2		GAMEIACIAG	AGUGGLIGIC							
HR3	TCTCCTGAA	AAACTACTOG	GOCAGACATC	AACAGGIGAA	TTTGAAGOCA	OGITICITAA	TOOCAAAAAA	ATAGTA.GTA	GAATATATCT	ACCACCOCAC
HR3	ACTICICAA	GAAGIAACGA	ANGAACTACC	AACCCACAAA	TITCAAGOCA	CTTTCTTCAA	TGGGAAAAAA	ATAG. AAGTA	CAATATACCT	ACAACCEACA
HR4	.ACTTOGCAA	QGACTTTTQG	GGCACCICAC	ATCACGAAAA	TTTCAAGCCC	GITTICIGAA	TGAGAAAGAA	ATAG. AAGTA	GAGTATACCT	ACCATOCAGA
HR5	TATTICICAA	GAACCACTOG	AGCAGCIACC	AACACGCAAA	CTICIAGICA	CITTICITAA	TAATAAAGAA	ATAG.AAGIG	GAATACATGT	ACGA
ORI								ATTAC ACCTTA	CANTACANCT	ATTACIOUSA
082	ATTICICCA	CAACING	GGCAACTATC	AALCAALAIC	ATTICERCEA	CITITIAGA	TYTAAAACAA	ATTOTA CTC	ACATATAAAT	ACCACINICA
Concentration		CAACTACTIC	COCA-CTATC	ANCACICTARA	TTTCAGOTA	CTTTTCTTAA	TYYZAAACAA	ATAG-AAGTA	GAATATA-CT	ACTACODOGA
0.1361505	American	Grenerou								
	101									200
BR1	GACAACTOCA	ATAAA	.TIGGATICT	TATCAATTIC	AAAGACTIGT	TACCAATACT	TGGTGGGATT	ATOCTACCAG	ATCATCAAGA	TAAATTOCAG
HR1	<u></u>									
HR2	TGTTCGCCAT	GGAGGG	CGIGITAA	AATTAATTTA	CAAGGATIGT	TACCAGIAIT	TUBUBBBAIC	TITITI	A. CAGAAIC	AAA.TIIGII
HDA	CACACTICAT	ALAGA	TYPE	ASTTAATTAA	GENGECTICT	TACAACAATT	TGATIGGATC	AAAT	TOGGTACOGT	AAACTTIGIC
HR5	GOGACITIOC	AAAGT	TOGIATTTTC	ATTIAATATA	CAAGCCTTGT	CACCAGIATT	CAATGOGATC	ATCCCACCCA	ATGAAAA	CAAATTTCAA
OR1	GAAACOCCGT	AACGAG	OGTCTCAA	ATTAATATAC	ATCICITATI	TACCCATATT	TOGTOGTOTC	AAGCTACCGA	ATAGGAGA	AAA.TTTCIG
OR2	GAGACCTOGT	AAAGAG	CETETTCA	AATCAATTTA	GAAGGATIGT	TGCCGGTAAT	TOGTOOGACC	ATGCCACCAG	ATGGGA.A	AAAGTTTTTG
OR3	GERGERGETT	ACTGAGGAGC	CITGIATIAA	GGICAATATA	GAAGACTTAT	TACCAGIACT	TOGTOGGACT	ATGOCACGAG	ATATGAAGCT	T
Consensus	GAGACGICGT	AAAGAG	-TOGIGITAA	AATIAATTIA	-AAGGCTIGT	TACCAGIATT	TGGIGGGAIC	AIGCIACC-G	AIG-GAAGA	AAAATTT-1G
	201									300
BR1	ACCATATICA	COCTACAACA	TTTTTTCAAA	AGAAACCTGA	AGGTACATCG	GIGTICAGGG	GGAATACATA	C.ATCCTGCC	ACAATCTOGG	COGTGATTOG
HRL	GA	GCTCCCAAGA	TTTGCTGAAA	AGTTACCTCA	ATOCACATGA	ATGTGACTOG	CAGOCOGAGA	CGAACCAAGA	CAAATCOCCC	ATCOGA
HR2	AGCACATCIG	ACCTGAAAGA	TITICIOGAA	AGTTACCTOG	ATTTACATGA	GIGIGCAGAT	GCAATATACA	C.ACCTIGCT	CCAATCTOCA	COCCCCTTC
HR3	AGAATGICIG	TOGTOCATTA	TATTTIAGCA	AAAAAACCCCGG	GTTTAGATAA	GIGIGGAGAI	GGAATATACA	C.ACCIGCC	CGAATCTOGA	COCAAATTIG
HR4	AACATATTCG	AGGTGGAAAT	TIATTIGAAA	CACGAGCIGG	AATTATATTC	GIGITIGGAA	GGAATGTTCA	C.ACCTIGCT	CGAATCTOGT	CTGGGATATG
HR5	ACCGTANAAT	CCATGACAGA	TCTTTTGAAA	CAAAATCIGG	ATTIACATCA	GIGITCAGAT	GGAAIGIACA	C.ACCAIGCA	CGAATTICGA	CONCEPTING
083	ACTINCIUS	ALGIUGAAG	TUTTTTTTT	CANACCTOC	ATTTACATCA	CITCHING AND	CONTINUTION	C ACCATIGCI	CAAATCTICCA	CAGOCATTIC
083	ASCALGIAIG	CONTRACTOR								
Consensus	AGCATATT-G	-OGTOGAAGA	TITTTTGAAA	A-AAACCTOG	ATTIACATCA	GIGIT-AGAT	GGAAT-TACA	C-ACCCTCC-	CCAATCTOCA	COG-GATTIG
	301									400
BRL UP1	GATICAGATT	CAGA	•••••	•••••	•••••	•••••	•••••	•••••	GGCGAAGITG	GALAICIGU
HRL HP2		•••••							CAATCACATA	AACTAATTOT
HR3	AA	••••••							GAAGGAATAC	GACGAATGCC
HR4	AA								Magnas.	GAATGOC
HR5	GACGAGAACG	ACCANCANCA	AGITTTTTAT	GAGIGIGIAA	GTGTAATGGA	CAACCTOCAC	TCAGAICIOS	ACCACCATT	GAAGCAGGAG	AAAAAATGTC
HR5 OR1	GAGGAGAACG	AGGAAGAAGA	AGTTTTTTAT	GAGIGIGIAA	GIGIAAIQGA	CAACCTOCAC	TCAGAICICS	ACCACCATT	GAAGGAGGAG GGTGGAGGAG	AAAAAATGTC GACACATGCC
HR5 OR1 OR2	GAGGAGAAOG 	AGGAAGAAGA	AGTTTTTTAT	GAGIGIGIAA	GTGIAATGGA	CAACCTOCAC	TCAGATCTOG	ACGAGGATTT	GAAGCAGCAG GCTGCAGCAG GCAGCAG	AAAAAATGTC GACACATGCC GATAGTTGCC
HR5 OR1 OR2 OR3	GAGGAGAAOG GA	AGGAAGAAGA	AGTTTTTAT	GAGTGTGTGAA	GTGTAATGGA	CAACCTOCAC		ACCACCATT	GAAGGAGGAG GGTGGAGGAG GGAGGAG	AAAAAATGTC GACACATGCC GATAGTTGCC
HR5 OR1 OR2 OR3 Consensus	GAGGAGAACG GA GA	AGGAAGAAGA	AGTTTTTTAT	GAGIGIGIAA	GTGTAATGGA	CAACCTOCAC		ACCACCATTT	GAACCACGAG GGTCGACGAG GGACGAG GGACGAG-AG	AAAAAATGTC GACACATGCC GATAGTTGCC GACAAATGCC
HR5 OR1 OR2 OR3 Consensus	GACCACAACG GA GA GA	AGGAAGAAGA	AGTTTTTTAT	GAGTIGTGTAA	GTGTAATGGA	CAAOCTOCAC		ACCACCATTT	GAAGCAGGAG GETGGAGGAG GGAGGAG GGAGGAG-AG	AAAAAATGTC GACACATGCC GATAGTTGCC GACAAATGCC 500
HR5 OR1 OR2 OR3 Consensus BR1	GAGGAGAAAG GA GA GA	AGGAAGAAGA	AGTTITITAT	GAGTGTGTAA	CIGIAATOGA			ACCACCATT TCG	GAACCACCAC GCTCCACCAC GCTCCACCAC GCACCAC-AC GCTCCAACCAC	AAAAAATGTC GACACATGCC GATAGTTGCC GACAAATGCC 500 TTTACGACGA
HR5 OR1 OR2 OR3 Consensus BR1 HR1	GAGGAGAAAG GA G		AGTITITIAT	ACCATGTGTAA	CIGIAATOGA	CAACCTCCAC	TACTCACGOCOC	ACCACCATT TCG	GAACCACCAC GCTCCACCAC GCACCAC GCACCAC-AC GCACCAC-AC CCTCCAACCAC CCTCCAACCAC	AAAAAATGTC GACACATGCC GATAGTTGCC GATAGTTGCC 500 TTTACGACGA GTTACGCCGA
HR5 OR1 OR3 Consensus BR1 HR1 HR2 HR2 HR3	GAGGAGAAAG GA GA GA GA GA CATACGACCA CACACITATCA CACACITATCA	AGGAAGAAGA TIACCATCAT CTITCATCAT TITCACATCAT	AGTITITIAT TICIGICAG TATIGICAG	ACCATGTGAA ACCATGTGAT ACCATGTGAT ACCATGTGAG AGCATGTGAG	GIGTAATGGA TGOGTGGTTT COOGTGGTTC TCATTGGTTC GGOGTGGTTC	CAACCTOCAC	TACAGATETES     TACTGACGGC     TACCGCAGC     TAACGCCAGC     TAACGCCAGC     TAACGCCAGC	ACCACCATTT TCG GATTTTACTT GATCATCCTA GATCATCCTA GATCATCCTA	GAACCACCAG GETCGAGGAG GGAGGAG GGAGGAG-AG CGTGAACCAG CGAGAACCAG CGAGAACCAG CGAGAACCAG CGAGAACCAG	AAAAAATGTC GACAATGCC GATAGTTGCC GACAAATGCC 500 TTTACGACGA GTTACGCCGA ATTACGCCGA
HR5 CR1 CR2 CR3 Consensus BR1 HR1 HR2 HR3 HR4	GACGAGAACC GA GA GA GA GA CATTICGACCA CACTACCA CATTACGATCA CATTACGATCA	ACCAACAACA TTACCATCAT CTTTCACAT TTTCACAT CTTTCACAT	AGTTITTAT	ACCATGTGAT ACCATGTGAT ACCATGTGAT ACCATGTGAT ACCATGTCAG ACCATGTCAG	GIGTAATGGA TGCGIGGTTT CGCGIGGTTC TCATIGGTTC GGCGIGGTTC TGCATGGTTC	CAACCTOCAC AAACATTATT AAGTATTACT AAGTATTACT GATTATTATT	TACTGACOCC TACTGACOCC TAACCCCACC TAACCCCACC TAACCACACC TAACCACACC	ACCACCATTT TOG 	GAACCACAGA GETCGACGAG GEACGAG GEACGAG COTGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG COAGAACGAG	AAAAAATGTC GACACATGCC GACAAATGCC GACAAATGCC 500 TTTACGACGA GTTACGACGA ATTACGCACC ATTACGCACC ATTACGCCC
HR5 CR1 CR2 CR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5	GACGAGAACG GA GA GA GA AO1 CATTICACCA CATTICACCA CATTICATCA CACACAATCA	ACCAACAACA TTACCATCAT CTTTCATCAT TTTCCACCAT CTTTCATCAT CTTCCACCAT CTTCCACCAT	AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TACIGIOCAG	ACCATGICAT ACCATGICAT ACCATGICAT ACCATGICAT ACCATGICAG ACCATGICAG ACCATGICAG ACCATGICAG	GIGIAATGA TGOGIGGITT GGOGIGGITT CATIGGTC GGOGIGGITC TGCATGGTC TGCATGGTC	CAACCTCCAC AAACATTATT AMGTATTACI AGGIGITACI GATTATTATT AAATATTATT AAATATTATT	TACTCACGO	ACCACCATTI TCG CATTITACTI GATCATCCTA GATCATCCTA GATCATCCTA AATTATCTTA AATTATCTTA	GAACCACAG GETCEACAG GETCEACAG GEACCAG-AG COTCEACCAG COT	AAAAAATGTC GACAATGCC GACAATGCC GACAAATGCC 500 TTTAGGACGA GTTAGGACGA ATTAGGCCGA ATTAGGCCGA ATTAGGCCGA ATTAGGCCGA
HR5 CR1 CR2 CR3 Consensus BR1 HR1 HR2 HR3 HR4 HR3 HR4 HR5 CR1	GACGAGAACC           GA	AGGAGAGA TTACCATCAT CTTTCATCAT TTTCCACCAT CTTTCATCAT CTTCCACCAT CTTCCACCAT	AGITITITAT TICTGIOCAG TOCTGIOCAG TATTGIOCAG TATTGIOCAG TATTGIOCAG TATTGIOCAG TATTGIOCAG	ACCATGTCAT ACCATGTCAT ACCATGTCAT ACCATGTAAA ATCATGTAAA ACCATGTTAT ACCATGTTAT	CICIDANICA CONTRACTOR TOCOLOGIUM CONTRACTOR TOCOLOGIUM TOCOLOGIUM TOCOLOGIUM TOCOLOGIUM TOCOLOGIUM	CAACCTCCAC AAACATTATT AAGTATTACT AGGIGTTACT GATTATTATT AAATATTATT AAATATTATT AAATAGTATT	TACTGACGC TACTGACGC TAACTGACG TAACTGACG TAACTGACG TATATGCTAC TATATGCTAC	ACCACCATTI TCC CATTITACITI CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA	GAACCACAG GETCGACAG GETCGACAG GEACCAG-AG CACCAG-AG CAGAACAG CAGAACAG CAGAACAG CAGAACAG CAGAACAG CAGAACAG CAGAACAG	AAAAAATGTC GACACATGCC GACAGTIGCC GACAAATGCC GACAAATGCC 500 TTTACGACGA ATTACGCCGA ATTACGCCGA ATTACGCCGA ATTACGCCGA ATTACGCACA ATTACGACAA
HRS ORI OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR1 OR1 OR2 OR1	GACGAGAACC           GA	AGGACAACA TTACCATCAT CTTTCATCAT CTTTCATCAT CTTTCATCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT	AGITITITAT TICTIGICAG TICTIGICAG TATTGICAG TATTGICAG TATTGICAG TATTGICAG TATTGICAG	ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT	CICITAATIGA CCCETCETTT CCCETCETTT CCCETCETTC CCCETCETTC CCCETCETTC TCCATCETTC TCCATCETTC TCCATCETTC TCCATCETTC	CAACCTCAC AAACATTATT AAGTATTACT GATTATTATT AAATATTATT AAATATTATT AAATATTAT	TACTGACOCC TACTGACOCC TACTGACOCC TACTGCACC TACTGCACC TACTGCACC TACCGCACC TACCGCACC TACCGCACC TACCGCACC TACCGCACC TACCACCC	ACCACGATTT TCG CATTTTACTT CATCATCCTA CATCATCATCT CATCATCATTA CATCACTATA TATCACTATA	GAACCACCAC GETCCACCAC GETCCACCAC GEACCAC-AC COTCAACCAC COCCACACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC COCCAACCAC	AAAAAATGTC GACAATGCC GACAATGCC GACAAATGCC 500 TTTACGACGA ATTACGCACA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGACAA ATTACGACAA ATTACGACAA
HR5 OR1 CR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR2 OR2 OR2	GACGAGAACG           GA           GATAGGAGA           GATAGGAGA           GATAGGAGA           GATAGGAGA           GATAGGAGA           GATAGGAGAGA           GATAGGAGAGAGA           GATAGGAGAGA           GATAGGAGAGAGA           GATAGGAGAGAGAGA           GATAGGAGAGAGAGA           GATAGGAGAGAGA           GATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	TPACCATCAT CITICATCAT CITICATCAT CITICATCAT CITICATCAT CITICATCAT CITICACCAT CITICACCAT CITICACCAT	AGITTTTTAT TICTGTOCAG TCTIGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG	ACCATGTGAT ACCATGTGAT ACCATGTGAT ACCATGTAA ATCATGTCAG ACCATGTAA ACCATGTTAA ACCAGTTAA	GIGIDAATGGA TGOGTGGTTT CGOGTGGTTC TCATTGGTTC TGCATGGTTC TGCATGGTTC TGCGTGGTTC TGCGTGGTTC	CAACCTCCAC AAACATTATT AAGTGTTACT ACGTGTTACT AAGTGTTACT AAATATTATT AAATATTATT AAATATTATT AAATATTAT	TACTGACGOC TACTGACGOC TACCGACG TACCGCAC TACCACAC TACACGCAC TACACGCAC TACCACGC	ACCACCATTI TCG GATTITACTI GATCATCTIA GATCATCTIA GATCATCTIA GATCACTATI GATCACTATI TATCACTATI	GAACCACCAC GETICGACCAC GETICGACCAC GEACCAC-AC COTIGAACGAC COAGAACGAC COAGAACGAC COAGAACGAC COAGAACGAC COAGAACGAC COAGAACGAC COGGAACGAC	AAAAAATGTC GACACATGCC GACAATGCC GACAAATGCC 500 GACAAATGCC 500 GACAAATGCC 30 GTTAGGCGCA ATTAGGCGCA ATTAGGCACA ATTAGGCACA ATTAGGCACA ATTAGGCACA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR1 OR1 OR2 OR3 Consensus	GACGAGAACG           GA           GA           401           CATTOGACGA           CATACGACA	TPACCATCAT CTTTACCATCAT CTTTACATCAT TTTOCACCAT CTTTCATCAT CTTCCACCAT CTTCCACCAT	AGITTTTAT TICTGICCAG TCCIGICCAG TATTGICCAG TATTGICCAG TATTGICCAG TATTGICCAG TATTGICCAG TATTGICCAG	ACCATIGTIGAA ACCATIGTIGAT ACCATIGTIGAT ACCATIGTIGAA ACCATIGTIGAA ACCATIGTIGAT ACCATIGTIGAT ACCATIGTIGAT ACCATIGTIGAT ACCATIGTIGAT	GIGIDAANGGA COSTIGGITT COSTIGGITC COSTIGGITC COSTIGGITC TICANGGITC TICANGGITC TICANGGITC TICANGGITC TICANGGITC	СААССТОСАС АААСАТТАТТ АСПАТТАТТА ССПАТТАТТА АСПАТТАТТА АААТАТТАТТ АААТАТТАТТ АААТАТТАТТ	TCAGAICIGG TACTGAGGG TAACGGAG TAACGGAG TAACGGAG TAACGGAG TAACGGAG TAACGGAG TAACGCAGG TAACCAGG	ACCACCATT TCC CATTTACTT CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCACCTATC CATCACCTAT CATCACCTATC	GAACACACA GETCGACAGA GETCGACAGA GEACGAG-AG CEACGAG-AG CEACGAGACAGA CEACGACAGAG CEACGACAGAG CEACGACAGAG CEACGACAGAG CEACGACAGAG CEACGACAGAG CEACGACAGAG	AAAAAATGTC GACACATGCC GATAGTTGCC GATAGTTGCC 500 TTTACGACAA GTTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCA ATTACGCCA ATTACGCCA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR1 OR2 OR2 OR2 OR2 OR2 OR2 OR2 OR2	GACCAGAACG GA GA 401 GATTICGACCA GATTICGACCA GATACCATICA CATTICGACCA GATACCATICA CATTICGACCA GATACCATICA CATTICGACCA CATACCATICA	TTACATCAT TTACATCAT CTTTCATCAT TTTCACAT CTTCATCAT CTTCCACAT CTTCCACAT CTTCCACAT	AGITTTTAT TICIGIOCAG TOURIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG	ACCATGICAT ACCATGICAT ACCATGITAC ACCATGITAC ACCATGITAC ACCATGITAT ACCATGITAT ACCATGITAT	GIGIDANIGA TGOGIGGITT CATIGGIC GOGIGGITC TGANGGITC TGCANGGITC TGCGIGGITT TGCGIGGITC	CAACCTCAC AAACATTATI AAGTATTACI ACGTGTTACI ACGTGTTACI AAGTATTATI AAATATTATI AAATAGTATI	TACTCACGCC TACTCACGCC TAACTCACGCAC TAACTCACAC TAACTCACAC TAACTCACAC TAACCACAC TAACCACGC TAACCACGC	ACCACGATT TCC CATTTACTT CATCATCCTA CATCATCCTA CATCATCCTA CATCATCCTA CATCACTCATC CATCACTATE CATCACTATE CATCACTATE CATCACTATE	GANGAGAGA GITGAGAGA GAGGAC-AG COTGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG COAGAAGAG	AAAAAATGTC GACAATGCC GATAGTTGCC GATAGTTGCC 500 TTTACGACA GTTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA ATTACGACA
HR5 QR1 QR2 Consensus BR1 HR1 HR2 HR3 HR4 HR5 QR2 QR3 Consensus SB1	GACCERCENCE           GA	AGGACAAGA TTACCATORT CTTTCATORT TTTOCACOAT CTTCCACACAT CTTCCACACAT CTTCCACACAT CTTCCACACAT CTTCCACACAT	AGITITITAT TICIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG	ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN ACCAUGUAN	TGOSTOSTIT COUSTOSTIT COUSTOSTIT COUSTOSTIT COUSTOSTIT TGORTOSTIT TGORTOSTIT TGORTOSTIT TGOSTOSTIT TGOSTOSTIT TGOSTOSTIT	CANCETOCAC AMACATTATT ANGTATTACI ACTIVITACI ACTIVITACI ANATOTITATI ANATOTITATI ANATATTATI ANATATTATI	TACTURACION TRACTORACION TRACTORACIONAL TRACTORACIONAL TRACTORACIONAL TRACTORACIONAL TRACCACIONAL TRACCACIONAL TRACCACIONAL TRACCACIONAL	ACCACGATT TO GATTITACTT GATCATCCTA GATCATCCTA GATCATCCTA GATCATCTTT TATCACTATT GATCATCATTTT TATCACTATTT GATCACTATTTT TATCACTATTTTTTTTTT	GANCHAGAG GETCHAGAG GETCHAGAG GETCHAGAG GETCHAGAG CETCHAGAG CEACHAG CEACHA	AAAAAATGTC GACACATGCC GATAGTTGCC GATAGTTGCC GTAGTGTGCC 500 TTTACGACGA ATTACGCCC ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACGCCCA ATTACCCCCA ATTACCCCCA ATTACCCCCA ATTACCCCCCA
HR5 OR1 OR2 Consensus BR1 HR2 HR3 HR4 HR5 OR1 OR2 OR3 Consensus BR1 BR1 HR4	GACCEAACC     GA     GA     GA     GA     GA     GA     GA     GA     AUDA     GATTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     CAPTOGAOCA     SO1     CCIPACIGNAGE     SO1     SO1     CCIPACIGNAGE     SO1	TEACATCAR TTACCATCAT CTTTCACCAT TTTCOCACCA CTTTCACAT CTTCOCACCA CTTCOCACCA CTTCOCACCA CTTCOCACCA CTTCOCACCA CTTCOCACCA CTTCOCACCA	AGITITITAT TICTGTOCAG TOCTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG	ACCAUGITGAL ACCAUGITGAL ACCAUGITGAL ACCAUGITGAL ACCAUGITAG ACCAUGITAG ACCAUGITAG ACCAUGITAG ACCAUGITAG ACCAUGITAG ACCAUGITAG	TGCSTGGTTT TGCSTGGTTT CGCSTGGTTC CGCSTGGTTC CGCSTGGTTC TGCATGGTC TGCATGGTC TGCATGGTC TGCSTGGTTC TGCSTGGTTC TGCSTGGTC ACATCTGGSC	CAACITCAC	TACTGROCOC TRACTOROCOC TRACTOROCOC TRACTOROCOC TRACTOROCOC TRACTOROCOC TRACTOROCOC TRACTOROCOC TRACCOCOC TRACCOCOC TRACCOCOC TRACCOCOC TRACCOCOC COCOCCENTRI	ACCACGATT 	GRACEACING GETCERAGES GETCERAGES GETCERAGES GETCERAGES CERCERAGES	Аналантотс GRCANIGCC GRCANIGCC GRCANIGCC GRCANIGCC STTENCSACS STO TTENCSACS ATTAICANS
HR5 (R1 (R2 (R3 Consensus HR1 HR2 HR2 HR2 (R4 (R4 HR2 (R4 (R4 (R4 HR2 (R4 (R4 (R4 (R4 (R4 (R4 (R4 (R4	GA     GA     GA     GA     GA     GA     A01     CATTGACA     CATTGACA     CATAGATA     CATAGATATA     CATAGATATA     CATAGATATA     CATAGATATA     CATAGATATA     CATAGATATATATATATATATATATATATATATATATATA	TEACATORY TEACATORY CTITICAION TITICAION CTITICAION CTITICAION CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH CTICONOCH	AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TACIGUAAGI CAGGAAACI	ACCATGIGATA ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT	TGGSTAGTOR TGGSTGGSTTT TGCATGGSTTC TGCATGGSTTC TGCATGGSTC TGCATGGSTC TGCATGGSTC TGCATGGSTC TGCATGGSTC TGCATGGSTC TGCGSTGGSTC ACATCTGGAC	CAACITOCAC AAACATTATT AAGTATTACT ACGTGTTACT AAGTATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT AAATATTACT	TORGATORO	ACCHOSATT ACCHOSATT CATTERNET GATCATOCER GATCATOCER GATCATOCER GATCATOCER GATCATOCER GATCATOCER GAT-AUGUE GAT-AUGUE GAT-AUGUE TGGCTACOTER	GINGRAGIA GEIGERAGIA GEIGERAGIA GEIGERAGIA GEIGERAGIA CERCINACIA C	AMAMANGTIC GACARIOSC GACARIOSC GARACTACC GACANATOCC SOO TTTACGACA ATTACTACA ATTACGACA ATTACA ATTACGACA ATTACA AT
HR5 (R1 QR2 QR3 Consensus BR1 HR2 HR2 HR3 HR4 HR5 QR2 QR3 Consensus BR1 HR2 HR3 HR4 HR5 QR2 QR3 R7 HR4 HR5 QR3 R7 HR4 HR7 HR7 HR7 HR7 HR7 HR7 HR7 HR7	GACCAACC     GACCAACC     GA     GA     GA     GA     GA     GA     GA     GA     GACAATCA     GACAATCA     GACAATCA     GATTICGACA	TEACATORI TEACATORI CTITOCIACIA CONTINUE CTITOCIACIA CTITOCIACIA CONTINUE CTITOCIACIA CTIT	AGITITITAT TTCTSTOCAG TCCTGTCCAG TATTGTCCAG TATTGTCCAG TATTGTCCAG TATTGTCCAG TATTGTCCAG TATTGTCCC-G TATTGTCCC-G TATTGTCCC-G CAGGACAT	ACCRIGICAL ACCRIGICAL		AAACATTATT AAACATTATT AACATTATTA AACATTATTAT AAATATTATTA AAATATTATTA AAATATTAT	TACTGACGO	ACCACGATT ACCACGATT GATCHTACTT GATCATCO GATCHTCGT GATCATCOTT GATCATCTT GATCACTTT GATCACTTT GATCACTTT GATCACTTT TGGCTACGTC TGGCTACGTC	GIACOLOGIA GETCOLAGIA GETCOLAGIA GETCOLAGIA GETACOLAGIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA COLGANACIA NOCTOLACIA	AMAMANGTIC GACARIOC GACARIOC GLAGHTICC GLAGHTICC GLAGHTICC SOO TTENCSACA STENCICA ATTENCICA ATTENCICA ATTENCICA ATTENCICA ATTENCICA ATTENCICA ATTENCICA ATTENCICA ATTACAGUE ATTA
HR5 OR1 OR2 OR3 Consensus HR1 HR1 HR1 HR4 HR4 OR2 OR3 OR2 OR3 Consensus RR1 HR1 HR2 HR2 HR2 HR2 HR2 HR2 HR2 HR2 HR2 HR2	GA     GA     GA     GA     GA     GA     A01     CAITCGACA     CAITCGACA     CAITCGACA     CAITCGACA     CAITCGACA     CAITCGACA     CAITCGACA     CAITCGACA     SO1     CCITCGACA     SO1     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA     CCITCGACA		AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACAAACC CACCAAACC CACCAAACC	ACCATGIGIA ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA CA	TGGETAATIGEA TGGETAATIGEA TGGETGGETTT TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC ACATCTTGGAG ACATCTTGGAG ACATCTTGGAG ACATCTTGGAG ACATCTTGGAG	AAACATTACT AAACATTACT AAACATTACT AACTACTACT CONTINUE AACTACTACT AAACATTACT AAACATTACT AAACATTACT AAACATTACT AAACATTACT CROCOCCERC ACCOCCERC	TACKARICTOC TACTGACGC TRACTACGC TRACTACGC TRACTACGC TRACCACGC TRACCACGC TRACCACGC TRACCACGC TRACCACGC TRACCACGC TRACCACGC COCCGGGUAT	ACCACGATT ACCACGATT CALIFICATION CALIFORMA CALIFORA	CHACTURES CETCERACES CETCERACES CETCERACES CERCES CERCERACES CERCE	AMAMANGTIC GACARIOOC GACARIOOC GACANATOCC SOO TTTINGSOCA STITACOCAC ATTRACACA ATTRACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTACACACACA ATTACACACACA ATTACACACAC
HR5 (R1 (R2 (R2 (R3 (R3 HR1 HR1 HR2 HR2 HR3 HR4 HR2 (R2 (R3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 HR3 (R4 (R4 HR3 (R4 HR3 (R4 (R4 (R4 (R4 (R4 (R4 (R4 (R4	GA     GA     GA     GA     401     CATICACCA     CATACACA     CATACACA     CATACACACA     CATACACACACACACACACACACACACACACACACA	TEACORTORY TEACORTORY CTITACORT CTITACORT CTITACATOR	AGITITITAT TICIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TACAGACAC CACGAAACIC CACGACACC CACGACACC TACAGICICAL	GAGIGIGIDA ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGITAS ACCARGITAS ACCARGITAS ACCARGITAS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS TRAIDACTS	TGGETAGTOR TGGETAGTOR TGGETGGETTT TGCATGGETTC TGCATGGETTC TGCATGGETT TGCATGGETT TGGETGGETT TGGETGGETT TGGETGGETT ACARTCTGGAC	CAACCTOCAC AAACATTACT ACTIVITACI ACTIVITACI ACTIVITACI ACTIVITACI ACTIVITACI AAATATTATT AAATATTATT AAATATTATT AAATATTATT AAATATTATT AAATATTATT AAATATTATT AAATATTATT AAATATTATT	TOKONICTOC TACTURACION TACCICACIÓN TACCICACIÓN TACCICACIÓN TACACIÓN TACACIÓN TACACIÓN TACACIÓN TACACIÓN TACACIÓN TACACIÓN COSCILIÓN COCOCIÓN TOCOCIÓN COCOCIÓN TOCOCIÓN COCOCIÓN COCOCIÓN TOCOCIÓN	ACCACGATT ACCACGATT CATTITICA CATTITICA CATCATICA	GNACHAGH GEIGENAGH GEIGENAGH GEIGENAGH GEIGENAGH CEIGEN	AMAMANGTIC         GACARNOCC           GACARNOCC         GARANTOCC           GARANTOCC         SOO           TTTACACACA         SOO           TTTACACACA         ATTACOCACA           ATTACOCACA         ATTACOCACA           ATTACOCACA         ATTACOCACA           ATTACOCACA         ATTACOCACA           ATTACOCACA         ATTACOCACA           ATTACOCACA         GOO           TAACACACACA         GOO           TAACACACACA         TAAACOCTACA           TAAACOCTACA         TAAACOCTACA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR1 OR2 OR3 BR1 HR4 HR5 CR1 HR2 OR3 CR2 OR3 CR2 OR3 CR2 OR3 CR2 CR2 CR3 HR4 HR5 CR2 CR3 HR4 HR5 CR2 CR3 HR4 HR5 CR2 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 HR4 HR5 CR3 CR3 CR4 CR4 CR4 CR4 CR4 CR4 CR4 CR4	GACCAACC     GACCAACC     GACCAACC     GACCAACCA     GACCAACCAACCA     GACCAACCAACCAACCAACCAACCAACCAACCAACC	ACCALCANCA TTACCALCAN CTITICALCAN TTTOCALCAN TTTOCALCAN TTTOCACAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCAN CTITICALCANCANC CTITICALCANCANC CTITICALCANCANCANC CTITICALCANCANCANCANCANCANCANCANCANCANCANCANCANC	AGITITITAT TICIGIOCAG TICIGIOCAG TICIGIOCAG TICIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACGACACC CACGACACC CACGACACC CACAAACC CACAAACC CACAAACC	GAGIGIGIDA ACCAUGIGA ACCAUGIGAT ACCAUGIGAT ACCAUGIGAT ACCAUGIGA ACCAUGITAC ACCAUGITAC ACCAUGITAC ACCAUGITAC ACCAUGITAC TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG TOALCAUCIG	GIGIAAROGA GIGIAAROGA TGGGIGGITT GGGGIGGITT GGGIGGITT GGAIGGITC GGAIGGITC GGAIGGITC GGAIGGITC GGAIGGITC ACATCIGGA ACATCIGGA ACATCIGGA ACACCIGGA ACACCIGGA ACACCIGGA	CAACCTOCAC AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAATTACTAC AAATTACTAC AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACT AAATTACTACTACTACTACTACTACTACTACTACTACTAC	TACTGAOST	ACCANCENTT ACCANCENT GAUTITUACITI GAUTAINGTH GAUTAINGTH GAUTAINGTH GAUTAINGTH GAUTAINGTH TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT TGGCTHACHT	GALCALCAL GETCORACE GETCORACE GETCORACE GETCORACE GETCORACE GETCORACE GETCORACE CERCENTRA CONSINCE CONSIN CONSINCE CONSIN CONSIN CONSIN CONSIN CONSIN CONSIN CONSIN C	ΑλΑλΑΝΤΟΤΟ GACARIOCO GACARIOCO GACARIOCO GACARIOCO CACARIOCO SOO TTEXCEGOCA SOO TTEXCEGOCA ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO ATTRACOCO CONTOCIONIO TARCHOCO CONTOCIONIO TARCOCO CONTOCIONIO TARACONCIO TARACOCIO CONTOCIONIO TARACOCIO CONTOCIONIO TARACOCIO CONTOCIONIO TARACOCIO CONTOCIONIO TARACOCIO CONTOCIONIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO TARACOCIO CONTOCIO TARACOCIO TARACOCIO CONTOCIO TARACOCIO TARACOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO CONTOCIO TARACOCIO
HR5 (R1 (R2 (R3 (R3 HR1 HR1 HR1 HR2 (R3 HR4 HR5 (R1 HR4 HR5 (R1 HR4 HR5 (R1 HR4 HR5 (R1 HR4 HR5 (R1 (R3 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 HR4 HR5 (R1 (R1 (R1 (R1 (R1 (R1 (R1 (R1	GA		AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACAAATOC CACAAATOC CACAAATOC CACAAATOC CACAAATOC CACAAATOC CACAAATOC CACAAATOC	ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL ACCRIGICAL TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN TOSATIACIÓN	TGGETAATIGEA TGGETGGETTT TGGETGGETTT TGGETGGETTT TGGETGGE	CAACCTOCAC AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT AAAGTATTACT	TACKARICTOC TACTGACGAC TRACTACGAC TRACTACGAC TRACTACGAC TRACCACGAC TRACCACGAC TRACCACGAC TRACCACGAC TRACCACGAC TRACCACGAC TRACCACGAC COCGACATAC COCGACGACATAC COCGACATAC	ACCARGATT ACCARGATT CALIFICATI CALIFICA	GRACIACIA GENERACIA GENERACIA GENERACIA GENERACIA GENERACIA COMUNICA COMUNI	AMAMANGTIC GACARIOOC GACARIOOC GACARIOOC GACARIOOC GACARIOC SOO TTTINGGACA STITACGACA ATTRACACA ATTRACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA CONTRACACACA CONTRACACACA CONTRACACACA TOCACACACA ATTRACACACACA CONTRACACACA CONTRACACACACA CONTRACACACACA CONTRACACACACA CONTRACACACACACACACACACACA CONTRACACACACACACACACACACACACACACACACACACAC
HR5 (R1 (R2 (R3 Consensus BR1 HR1 HR2 HR3 HR4 HR3 CR2 (R3 Consensus BR1 HR1 HR2 (R4 HR3 Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus Consensus	GACCARCE     GACCARCEAR     GACCARCEARA      GACCARCEARA      GACCARCEARA	ACCARCANCA           TTACCATCAT           CTTTACCATCAT           TTTACCATCAT           CTTTCATCAT           CTTTCATCAT           CTTCATCATCAT           CGTGGTGTGCACAT           CGTGGTGTGCACAT           CGTGGTGTGCACAT           CGTGGTGTGCACAT           CGTGGTGTGCACAT           CGTGGTGTGCACAT	AGITITITAT TICIGIOCAG TOCIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG TALTIGIOCAG CACGAAACC CACGAAACC CACAAACCAG CACAAACCAG	ACCARGINATION AC	TGGGTQGTHT GGGTQGTTT GGGTQGTTT TGATTQGTTC GGGTQGTTT TGATQGTTC TGATQGTTC TGGTQGTTT TGGTQGTTT TGGTQGTTT TGGTQGTCTTC ACATCTQGAC ACACCQGGA ACACCQGGA ACACCQGGA ACACCQGGA ACACCQGGA	CAACCTOCAC AAACATTACT ACGACTACT ACGACTACT ACGACTACT ACGACTACT ACGACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT ACGACCTAC ACGACCTAC ACGACCTAC ACGACCTAC ACGACCTAC	TORGATOR	ACCACGATT ACCACGATT CATTTACT GATCATCAC GATCATCAC GATCATCAC GATCATCAC GATCATCAC GATCATCAC TACCACACAC GATCACTAC GAT-ALGET TACCACACAC TACCACACACACACACACACACACACACACACACACACA	GNACHAGN GEICENAGH GEICENAGH GEICENAGH GEICENAGH GEICENAGH CEIGEN	AMAMANGTIC GACARITOCC GACARATOCC GACANATOCC SOO TTTACACOCA ATTRACOCAC ATTRACO
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 OR1 OR2 OR3 HR1 HR5 OR1 HR2 OR3 Consensus Consensus Consensus	GACCHARCE GACCHARCE GACCHARCE GACCHARCE GACARTAC CATACENCA CATACEN	ACCALCANCA TTACCALCAT CTITICALCAL TTTOCALCAT TTTOCALCAL CTITICAL CTITICALCAL CTITICAL CTITICAL CTITICAL CTITICAL CTITICAL CTITICA	AGITITITAT TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG TICTGTOCAG CAGGAG CAGGAG CAGGAGAG CAGGAGAG CAGGAGAG CAGGAGAG CAGGAGAG CAGGAG C CAGGAG CAGGAG CAGGAG CAGGAG CAGGAG CAGGAG C CAGA	ACCATGIGIA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIA ACCATGIA TGATACTG TGATACTG TGATACTG TGATACTG TGATACTG TGATACTG	TGGGTGGTATIGGA GGGGTGGTTT TGGGTGGTTT TGGGTGGTTT TGGATGGGTTT TGGATGGGTTC TGGATGGGTTC TGGATGGGTTC TGGATGGGTTC TGGATGGTTC TGGGTGGTTC ACACTGGGG ACACTGGGG ACACCGGGA ACACCGGGA ACACCGGGA	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAATAGTACTACT AAATAGTACTACT AAATAGTACTACT AAATAGTACTACT AAATAGTACTAC AGGACOCTAC ACCOGCCTACA ACCOGCCTACA	TACTGACOM TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN TRACTICACIÓN COSSIGNATIAN COSSIGNATIAN COSSIGNATIAN COSSIGNATIAN COSSIGNATIAN	ACCACGATT ACCACGATT CALITIACIT CALITIACIT CALITIACIT CALITIACIT ANTINATIT ANTINATIT CALITIACIT TALCACITACI CALITIACIT TALCACITACI CALITIACIT TALCACITACI CALITIACIT TALCACITACI TALCACI	GRACENCIE GETORACIÓN GETORACIÓN GETORACIÓN GETORACIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN CERCIANCIÓN ACOLLINACIÓN CONCINACIÓN	AMAMANGTIC GACARITOCC GACARITOCC GACARITOCC GACARITOCC GITAGTOCA STTACSACA STTACSACA ATTACACA ATTACSACACA ATTACSACA ATTACSACACA ATTACACA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR4 HR5 OR2 OR3 Consensus Consensus Consensus	GACCAACC     GACCAACC     GACCAACC     GACCAACCAACCAACCAACCAACCAACCAACCAACC	TTACCALCANCA CTITICALCAL CTITICAL	AGTITITIAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACAAATCC CACAAATCC CACAAATCC CACAAATCC CACAAATCC CACAAATCC CACAAATCC CACAAATCC CACAAATCC	ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION ACCRIGICATION TEATCACCTO TEATCACCTO TEATCACCTO TEATCACCTO TEATCACCTO TEATCACCTO TEATCACCTO TEATCACCTO	TGGETAGTGA TGGETGGETTT TGGETGGETTT TGGETGGETTT TGGETGGE	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAATATTACT	ТОКСКИСТСК ТАСТСЯКОСС ТРАСТСЯКОСС ТРАСТСЯКОСС ТРАССИСКИ ТРОССИСКИ ТРОССИСИ ТРОССИСКИ ТРОССИСИ ТРОССИСИ ТРОССИ ТРОССИСИ ТРОССИСИ ТРОССИСИ ТРОССИ ТРОССИ ТРОССИСИ	ACCACGATT ACCACGATT CALIFICATI CALIFICA	GRACIACIA GENERACIA GENERACIA GENERACIA GENERACIA GENERACIA CERCIANO CERCIN	AMAMANGTIC GACARIOGC GACARIOGC GUNATIGOC 500 TTTINGGOCA GUTAGOCOA ATTROGACA ATTROGACA ATTROGACA ATTROGACA ATTROGACA ATTROGACA ATTROGACA ATTROCACACA ATTROCACA ATTROCACA ATTROCACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACACA ATTROCACACACACACACACACACACACACACACACACACACA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 HR3 Consensus BR1 HR4 HR5 OR1 OR2 OR3 Consensus BR1 HR4 HR5 OR1 OR2 OR3 OR1 OR1 OR2 OR3 OR3 OR3 OR3 OR3 OR3 OR3 OR3	GACCAACC     GACCAACC     GACCAACC     GACCAACA     GA     GACCAACA     GACCAACAA     GACCAACAACA     GACCAACAACAA     GACCAACAACAA     GACCAACAACAA     GACCAACAACAA     GACCAACAACAA     GACCAACAACAACAACAACAACAACAACAACAACAACAAC	ACCARANCA TTACCAICAT TTACCAICAT TTTCAICAT TTTCAICAT TTTCAICAT TTTCAICAT TTTCAICAT TTTCAICAT TTTCAICAT CTTCAICAT	AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACGAAACC CACGAAACC CACGAAACC CACGAAACC CACGAAACC	ACCARGIGATIA ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT TGATCACCIG TGATCACCIG TGATCACCIG TGATCACCIG TGATCACCIG TGATCACCIG	TGGGTQGTTT TGGGTQGTTT TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTC TGCATGTC TGCATGGTC	CAACCTOCAC AAACATTACT ACTIVITACI ACTIVITACI ACTIVITACI ACTIVITACI AAATATTACT	TORGATOR	ACCACGATT ACCACGATT CATTTACT GATCATCAC GATCATCAC GATCATCAC GATCATCAC GATCATCAC GATCATTACT TACCACTACA CATTACTTAC TACCACTACACA TACCACTACA TACCACTACA TACCACTACACA TACCACTACACA TACCACTACACACACA TACCACTACACACACACACACACACACACACACACACAC	GNACHAGIG GEICGNAGH GEICGNAGH GEICGNAGH GEICGNAGH GEICGNAGH CEIGENAGHAG CEIGENAGHAG CEIGENAGHAG CEIGENAGHAG CEIGENAGHAG CEIGENAGHAG CEIGENAGH CEIG	AMAMANGTIC GACARITOCC GACARITOCC GACARATOCC SOO TTTACACOCA ATTRACACOCACACOCA ATTRACACOCACACOCAC ATTRACACOCACACOCACACOCACOCACOCACOCACOCACOCA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 OR3 OR1 OR2 OR2 OR3 BR1 HR2 OR1 OR2 OR3 Consensus BR1 HR2 OR2 OR3 OR2 OR3 BR1 HR3 HR3 HR3 Consensus BR1 HR3 HR3 HR3 Consensus BR1 HR3 HR3 HR3 HR3 HR3 HR3 HR3 HR3	GACCARACC GACCARACC GACCARACCARACCARACCA	TTACCALARA TTACCALCAT CTTTCALCAL CTTTCALCAL TTTCCALCAL CTTTCALCAL CTTTCALCAL CTTCCA	AGITTITIAT TICTGTOCAG TOCTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG CAGGAGACT CAGGAGACT CAGGAGACC CAGGAGACC CAGGAGACC CAGGAGACC CAGGAGACC CAGGAGACC	ACCATGIGIA ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGAT TGATCACTG TGATCACTG TGATCACTG TGATCACTG TGATCACTG TGATCACTG	TGCGTGGTATTGGA GGCGGGTTT TGCGTGGTTT TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCGTGGTTC ACATCTGGAC ACATCTGGAC ACATCTGGAC ACATCTGGAC ACATCTGGAC ACATCTGGAC ACATCTGGAC ACATCGGAC ACATCGGAC ACATCGGAC ACATCGGAC	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAATATTACT	TACTGACGA TACTGACGCC TRACTGACGCC TRACTGACG TRACCACCG TRACCACCG TRACCACCG TRACCACCG TRACCACCG TRACCACCG CCGGGAATAT CCCCGGAATAT CCCCGGAATAT CCCCGGAATAT	ACCACGATT ACCACGATT GAUTATCIAC GAUCATCOTT CAUCATOR CAUCATOR CAUTOROTT AATTANTOTT TAUCACTACI CAUTOROTT TAUCACTACI CAUTACOTTACI CAUTACOTTACI CAUTACOTTACI CAUTACOTTACI CAUTACOTTACI CAUTACOTTACI CAUTACOTTACI CONTRACT TOSCTACI TOSC	GRACHAGIC GEIGERA-L GEIGERA-L GEIGERA-L GEIGERA-L GEIGERA-L GEIGERA-L CEIGERANGER CEIGER CE	AMAMANGTIC GACARIOCC GACARIOCC GACARIOCC GACARIOCC GIUMOTTOCC SOO TTIPAGACA ATTROGACA ATTRACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA ATTRACACACA TAACARITC
HR5 GR1 GR2 GR3 Consensus BR1 HR1 HR2 HR3 HR4 HR5 GR1 GR2 GR3 Consensus Consensus Consensus BR1 HR2 GR3 CR2 GR3 CR2 GR3 CR2 GR3 CR2 GR3 CR2 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR3 CR2 CR3 CR2 CR3 CR3 CR3 CR3 CR3 CR3 CR3 CR3	GACCAACC GACCAACC GACCAACCAACCAACCAACCAA	ACCARCANCA           TTACCATCAT           TTTACCATCAT           TTTTCATCAT           TTTTCATCAT           TTTTCATCAT           CTTTCATCAT           CTTCATCATCAT           CTTCANACAT           CTTCANCAT           CTTCANACAT           CTTCANACAT           CTTCANCAT           CTTCANCAT           CTTCANCAT           CTTCANCAT           CTTCANCAT           CTTCANCAT           CTTCANCATCAT           CTTCANACCATCAT           GANAC           CENACCAT	AGITITITAT TICIGIOCAG TOCIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACAAATOC CACGAAACC CACAAATOC CACAAATOC CACAAACCC	ACCATGIGIA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI TGATCACCI	TGGETAGTOR TGGETGGETT TGGETGGETT TGGETGGETTC TGGETGGETTC TGGETGGETTC TGGETGGETTC TGGETGGETTC TGGETGGETTC ACATCTTGGEG ACATCTTGGEG ACATCTTGGEG ACATCTTGGEG ACATCTTGGEG ACATCTTGGEG ACATCTGGEG ACATCGGEG ACATCGGEG ACATCGGEG ACATCGGEG	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT ACTIGUTACT GRITTATTACT AAATGTTACT AAATGTTACT AAATGTTACT AAATGTTACT AAATGTTACT AAATGTTACT AAATGCTTAC ACCOCCTAC	ТОКСКИСТСК ТРАСТСЯКОВС ТРАСТСЯКОВС ТРАСТСЯКОВС ТРАССИСКО ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВС ТРАССИСКОВСКОВС ТРАССИСКОВСКОВС ТРАССИСКОВСКОВСКОВСКОВСКОВСКОВСКОВСКОВСКОВСКОВ	ACCALCATT ACCALCATT CALITYTACTT CALITYTACTT CALICALCATT CALICALCATT CALITALICAT CALITALICAT CALITALICAT CALITALICAT TACACIMAN CALITALICAT TACACIMAN TOCCTACOTT TOCCTACOTT TOCCTACOTT TOCCTACOTT TOCCTACOTT TOCCTACOTT TOCCTACOTT TOCCTACOTT	GNG2AGAG GETG2AGAG GETG2AGAG GEGG2AG-AG GEGG2AG-AG GEGG2AGAG GEGGAAGAG GEGGAAGAG CGGAAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAAG CGGAAGAG CGGAAGAG CGGAAGAG CGGAG C	AMAMANGTIC GACARIOOC GACARIOOC GACARIOC SOO TTTINGGACA SOO TTTINGGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRAGACA ATTRACACION ATTRACACION TAACARITA CATOROGICA TAACARITA TCACOCOTT TCACIONITA TCACOCOTT TCACIONITA
HR5 GR1 GR2 GR3 Consensus BR1 HR1 HR2 HR3 GR1 GR2 Consensus Consensus Consensus BR1 HR4 HR5 GR1 GR2 GR3 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR4 HR5 GR1 HR5 HR4 HR5 GR1 HR5 HR4 HR5 HR4 HR5 GR1 HR5 HR4 HR5 HR5 HR5 HR4 HR5 HR5 HR5 HR5 HR5 HR5 HR5 HR5	GALERAACC GALERAACC GALERAACCA GALERAACCA GALERAACCA GALERAACCA GALERAACCA CA	ACCALCANCA TTACCAICAT TTTCACAICAT TTTCACAICAT TTTCACAICAT TTTCACAICAT TTTCACAICAT TTTCACAICAT TTTCACAICAT CTTCACAICAT CTTCACAICAT CTTCACAICAT CTTCACAICAT CTTCAAAAAC CTTCAAAAAC CTTCAAAAAC CTTCAAAAAC CTTCAAAAC CTTCAAAAC	AGTITITIAT TICTGTOCAG TICTGTOCAG TICTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG TATTGTOCAG CAGGAAACC CAGGAAACC CAGGAAACC CAGGAAACC CAGGAAACC CAGGAAACC CAGGAAACC CAGGAAACC	ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGIGAT ACCARGITAG ACCARGITAG ACCARGITAG CARGIGAT CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG CARDACCIG	TGGGTGGTTT TGGGTGGTTT TGATTGGTC TGATTGGTC TGATTGGTC TGATGGTC TGGATGGTTT TGGATGGTTT TGGATGGTTT TGGATGGTTT TGGATGGTTT TGGATGGTTT TGGATGGTTT TGGATGGTTC ACACTGGA ACACTGA	AAACATTACT AAACATTACT ACGACTACT ACGACTACT ACGACTACT ACGACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT AAATACTACT ACCOCCTAC ACCOCCTAC ACCOCCTAC ACCOCCTAC	TOKONO CONTRACTOR CONT	ACCACGATT ACCACGATT CATTITICA GALCALCIT GALCALCIT GALCALCIT GALCALCIT CATTICAL	GNACHAGIG GEICENAGH GEICENAGH GEICENAGH GEICENAGH GEICENAGH CEIGEN	AMAMANGTIC GACARIOCC GACARIOCC GACARATOCC SOO TTTRACACAR ATTRACACAR ATTRACACAR ATTRACACAR ATTRACACAR ATTRACAR A
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR2 OR3 OR1 OR2 OR3 OR2 OR3 Consensus BR1 HR2 OR2 OR3 CR2 OR3 CR2 OR3 CR2 OR3 CR3 CR2 OR3 OR1 OR2 OR3 OR2 OR3 OR3 OR3 OR3 OR3 OR3 OR3 OR3	GACCARACC GACCARACC GACCARACCARACCARACCA	ACCARANCA ACCARANCA CTTTACCATCAT CTTTACATCAT CTTTCATCAT CTTTCATCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACCAT CTTCCACAT C	AGITITITAT TICTGIOCAG TOCIGIOCAG TATIGICAG TATIGICAG TATIGICAG TATIGICAG TATIGICAG TATIGICAG TATIGICAG TATIGICAG CACGAAACC CACGAAACC CACGAAACC CACGAAACC CACGAAACC	ACCATGIGIA ACCATGIGAT ACCATGIGAT ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA TGATCACCG TGATCACCG TGATCACCG TGATCACCG TGATCACCG TGATCACCG	TGCSTAGTATICGA GGGTATICGA GGGGTGGTTC TCATTGGTTC GGGGTGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC TGCATGGTTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC ACATCTGGTC	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT AAGTATTACT AAATATTACT	TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION TRACTICACION CONSIGNATIVI CONSIGNATIVI CONSIGNATIVI CONSIGNATIVI CONSIGNATIVI	ACCACGATT CALIFICATI CALIFICATI CALIFICATI CALIFICATI CALIFICATI ANTIVITY ANTIVITY CALIFICATI	GRACHAGE GETORAGEG GETORAGEG GERGRA-HO GERGRACEA GERGRACEA GERGRACEA GERGRACEA GERGRACEA GERGRACEA CONSTRUCTION CONSTRUCTION ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA ACCELLANCEA	AMAMANGTIC GACARIOOC GACARIOOC GACARIOOC GIUMOTTOC SOO TTTMCGOCA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA ATTACAGACA TACAGACACA TACAGACA TACAGACACA TACAGACACA TACAGACACACA TACAGACA
HR5 OR1 OR2 OR3 Consensus BR1 HR1 HR1 HR2 OR2 OR3 Consensus Consensus Consensus BR1 HR2 HR3 CR2 OR3 CR2 OR3 CR2 OR3 CR3 CR2 OR3 CR2 CR3 CR2 CR3 CR2 CR3 CR2 CR3 CR3 CR2 CR3 CR3 CR3 CR2 CR3 CR3 CR3 CR3 CR3 CR3 CR3 CR3	GACCARACC GALLARACCARACC GALLARACCARAC GALLARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC CARACCARAC SOI CENTROLOGIC CARACCARAC CARACCAR	ACCARCANCA           TTACCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCAT           CTTTCATCATCAT           CTTTCATCATCAT           CTTTCATCAT           CTTTCATCAT           CTTCATCATCAT           CATCATTA           CANAC           CGNACC	AGITITITAT TICIGIOCAG TICIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG TATIGIOCAG CACAAAICC CACAAAICC CACAAAICC CACAAACCC	ACCATGIGAT ACCATGIGAT ACCATGIGAT ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA ACCATGIGA TGATCACCTG TGATCACCTG TGATCACCTG TGATCACCTG TGATCACCTG TGATCACCTG TGATCACCTG	TGGETAATIGEA GGETAATIGEA GGGETGGETTT TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC TGGATGGETTC ACATCTGGAG ANCACTGGAG ANCACTGGAG ANCACTGGAG ACACTGGAG ACACTGGAG ACACTGGAG ACACTGGAG ACACTGGAG	AAACATTACT AAACATTACT AAGTATTACT AAGTATTACT ACGIGTIACI GRITALITACI AAATAGTIACI AAATAGTIACI AAATAGTIACI AAATAGTIACI AAATAGTIACI AAATAGTIACI AAATAGTIACI ACOGOCITAC ACOGOCITAC	TOKENICTO           TACTERCOC           TACTERCOC           TACTERCOC           TACTERCOC           TACTERCOC           TACATERCE           TACATERCE           TACATERCE           TACATERCE           TACATERCE           TACATERCE           TACATERCE           CONSTRUCTOR           CONSTRUCTOR <td>ACCACGATT ACCACGATT CATTTACTT CATCATCA CATCATCA CATCATCAT CATCATCATT CATCATCATTA CATCATCATTA CATCATCATA CATCATCATA CATCACTACATATA CATCACTACATATA CATCACTACATATA CATCACTACATATA CATCACTACATATATATATATATATATATATATATATAT</td> <td>GNG2AG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 CERCSAG</td> <td>AMAMANGTIC GACARIOOC GACARIOOC GACARIOOC GUNATIGOC SOO TTITAGOCACA SOO TTITAGOCACA ATTROCACACA ATTROCACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACACA ATTROCACACACACACACACACACACACACACACACACACACA</td>	ACCACGATT ACCACGATT CATTTACTT CATCATCA CATCATCA CATCATCAT CATCATCATT CATCATCATTA CATCATCATTA CATCATCATA CATCATCATA CATCACTACATATA CATCACTACATATA CATCACTACATATA CATCACTACATATA CATCACTACATATATATATATATATATATATATATATAT	GNG2AG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 GETCSAG36 CERCSAG	AMAMANGTIC GACARIOOC GACARIOOC GACARIOOC GUNATIGOC SOO TTITAGOCACA SOO TTITAGOCACA ATTROCACACA ATTROCACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACA ATTROCACACACA ATTROCACACACACACACACACACACACACACACACACACACA

CR3 ..... Consensus GGTTT-TCTC G-AAG----

FIG. 6. Optimal alignment of the nine sequenced repeat elements (partial and complete) from pSH-B, -H, and  $-O^1$ . The locations of the five repeats from pSH-H (HR1, nt 1 through 297; HR2, nt 298 through 822; HR3, nt 823 through 1351; HR4, nt 1352 through 1869; HR5, nt 1870 through 2462), the three repeats from pSH-O<sup>1</sup> (OR1, nt 1 through 442; OR2, nt 443 through 964; OR3, nt 965 through 1153), and the single repeat of pSH-B (BR1, nt 44 through 587) were determined from the dot matrix homology plots in Fig. 5. Gaps (periods) were inserted into the sequences to obtain a best fit. Dashes indicate that no consensus nucleotide was found for that specific location. The plurality used for the consensus sequence was 3.

determined by a comparison of the nine sequenced repeat elements (Fig. 6) did not reveal any distinguishing features such as the internal inverted or direct repeats which have been shown to exist in other eucaryotic repeated elements (9). The consensus sequence may be similar to an ancestral DNA that was the progenitor of the family of repeats we identified in this study.

The sequence data from  $pSH-O^1$ , -H, and -B also revealed six ORFs within the sequences of the repeated elements (Fig. 7). Comparison of the computer-generated predicted

translation products of each of the six ORFs showed amino acid homology. Recently, G. W. Blissard, O. P. Smith, and M. D. Summers (Virology, in press) identified two CsV genes expressed in parasitized *H. virescens* that share significant nucleotide and amino acid sequence homology and appear to represent two members of a larger, related viral gene family. The nucleotide sequences of these two related genes show no homology to the repeated elements identified in this study. The ORFs identified in the 540-bp repeat elements of pSH-B, -H, and  $-O^1$  may represent another





FIG. 7. Summary schematic showing the location of the 540-bp repeat elements and the ORFs of the sequenced regions of p5H-H (A), pSH-O<sup>1</sup> (B), and pSH-B (C). The locations of the direct repeat elements are shown by the arrows. The sizes of each ORF are as follows: 903 nt (H ORF 1), 996 nt (H ORF 2), 440 nt (H ORF 3), >853 nt (O ORF 1), 291 nt (B ORF 1), and 375 nt (B ORF 2). The open end of O<sup>1</sup> ORF 1 indicates that the end of the ORF is beyond the sequenced region.

example of a family of related CsV genes. Studies are currently in progress to investigate which, if any, of the six ORFs identified on the cloned SH DNAs B, H, or  $O^1$  are actively transcribed into CsV mRNAs in either *C. sonorensis* or parasitized *H. virescens*. A 3.2-kb mRNA expressed in parasitized *H. virescens* which is homologous to and is believed to be transcribed from pSH-H has been previously reported (2, 3). This mRNA would be large enough to code for any or all of the three ORFs we identified on pSH-H.

The organization of repeat elements in tandem arrays of various lengths on the multiple SH DNAs of the CsV genome could provide a substrate for a large number of intraand intermolecular homologous recombinational events, including equal and unequal crossover (10, 17). Multiple recombinational events combined with sequence divergence could result in a set of DNAs that would have very complex sequence relationships similar to what we presently observe for the CsV genome. It has been suggested that recombinational events between tandem arrays of repeats may provide a method of gene duplication and mutation of eucaryotic genes (10, 30). In the C. sonorensis-H. virescens parasitehost system, multiple recombinational events in CsV could be of evolutionary advantage for the parasitic wasp. Such changes may be related to the ability of the parasitic wasp to successfully parasitize its lepidopteran hosts that may be evolving new or altered defensive mechanisms.

In tissues of male and female C. sonorensis, sequences homologous to CsV DNA are found in nonviral forms which may be indicative of CsV integration in the wasp genome (13). Repeat elements in other viral and eucaryotic systems have been shown to be involved in genomic integration events (5, 21). If the replicative cycle of CsV does involve integrated viral DNAs, the 540-bp repeated elements identified in this study may also play a similar role.

In summary, the data in this study show that the multipartite CsV genome is composed of interrelated SH DNAs, indicative of a complex evolutionary history. Shared sequences between SH DNAs include both closely related

and diverged sequences; an example of the latter being the imperfectly conserved repeated elements that are homologous to most of the SH DNAs in the CsV genome. The presence of ORFs within the sequenced repeated elements suggests that they may represent a group of related viral genes. Alternatively, the repeated elements may be sites of homologous recombination between the SH DNA molecules or the putative integrated forms of viral DNA. Transcriptional analysis of the repeat sequences in either *C. sonorensis* or parasitized *H. virescens* may reveal some of the functional aspects of this family of repeated elements.

#### ACKNOWLEDGMENTS

The authors thank Gary W. Blissard for very helpful discussions and critical review of this manuscript and Burt Beames for editorial comments. We also thank Nelida Angulo for excellent technical assistance in the rearing of insects.

This work was supported by National Science Foundation grant PCM-8319002 and by Texas Agricultural Experiment Station Project 6313. D.A.T. is supported by the Research Scientist Trainee Program of Agriculture Canada (Vancouver).

### LITERATURE CITED

- Beltz, G. A., K. A. Jacobs, T. H. Eickbush, P. T. Cherbas, and F. C. Kafatos. 1983. Isolation of multigene families and determination of homologies by filter hybridization methods. Methods Enzymol. 100:266–285.
- Blissard, G. W., J. G. W. Fleming, S. B. Vinson, and M. D. Summers. 1986. *Campoletis sonorensis* virus: expression in *Heliothis virescens* and identification of expressed sequences. J. Insect Physiol. 32:351–359.
- 3. Blissard, G. W., S. B. Vinson, and M. D. Summers. 1986. Identification, mapping, and in vitro translation of *Campoletis* sonorensis virus mRNAs from parasitized *Heliothis virescens* larvae. J. Virol. 57:318-327.
- Brown, F. 1986. The classification and nomenclature of viruses: summary of results of meetings of the International Committee on Taxonomy of Viruses in Sendai, September 1984. Intervirology 25:141-143.
- 5. Carlos, P. M., and J. H. Miller. 1980. Transposable elements.

Cell 20:589-595.

- 6. Cochran, M. A., and P. Faulkner. 1983. Location of homologous sequences interspersed at five regions in the baculovirus AcMNPV genome. J. Virol. 45:961–970.
- Dale, R. M., B. A. McClure, and J. P. Houchins. 1985. A rapid single stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid 13:31-40.
- 7a. Davies, D. H., M. R. Strand, and S. B. Vinson. 1987. Changes in differential haemocyte count and in vitro behaviour of plasmatocytes from host *Heliothis virescens* caused by *Campoletis* sonorensis polydnavirus. J. Insect Physiol. 33:143–153.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
- 9. Doolittle, W. F. 1985. The evolutionary significance of middle repetitive DNAs, p. 443–487. *In* T. Cavalier-Smith (ed.), The evolution of genome size. John Wiley & Sons, Inc., New York.
- 9a.Dover, B. A., D. H. Davies, M. R. Strand, R. S. Gray, L. L. Keeley, and S. B. Vinson. 1987. Ecdysteroid-titre reduction and developmental arrest of last-instar *Heliothis virescens* larvae by calyx fluid from the parasitoid *Campoletis sonorensis*. J. Insect Physiol. 33:333–338.
- 10. Dyson, P., and D. Sherratt. 1985. Molecular mechanisms of duplication, deletion, and transposition, p. 353–395. *In* T. Cavalier-Smith (ed.), The evolution of genome size. John Wiley & Sons, Inc., New York.
- Edson, K. M., S. B. Vinson, D. B. Stoltz, and M. D. Summers. 1981. Virus in a parasitoid wasp: suppression of the cellular immune response in the parasitoid's host. Science 211:582-583.
- Fleming, J. G. W., G. W. Blissard, M. D. Summers, and S. B. Vinson. 1983. Expression of *Campoletis sonorensis* virus in the parasitized host, *Heliothis virescens*. J. Virol. 48:74–78.
- 13. Fleming, J. G. W., and M. D. Summers. 1986. Campoletis sonorensis endoparasitic wasps contain forms of C. sonorensis virus DNA suggestive of integrated and extrachromosomal polydnavirus DNAs. J. Virol. 57:552–562.
- Galinski, M. S., V. H. Stanik, G. F. Rohrmann, and G. S. Beaudreau. 1983. Comparison of sequence diversity in several cytoplasmic polyhedrosis viruses. Virology 130:372–380.
- Howley, P. M., M. A. Isarael, M. Law, and M. A. Martin. 1979. A rapid method for detecting and mapping homology between heterologous DNAs. J. Biol. Chem. 254:4876–4883.
- 16. Krell, P. J., M. D. Summers, and S. B. Vinson. 1982. Virus with a multipartite superhelical DNA genome from the ichneumonid

parasitoid Campoletis sonorensis J. Virol. 43:859-870.

- 17. Lewin, B. 1980. Gene expression, vol. 2, p. 503-569. John Wiley & Sons, Inc., New York.
- Maizel, J. V., and R. P. Lenk. 1981. Enhanced matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA 78:7665-7669.
- 19. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Palmer, J. D., and C. R. Shields. 1984. Tripartite structure of the Brassica campestris mitochondrial genome. Nature (London) 307:437-440.
- 21. Panganiban, A. T. 1985. Retroviral DNA integration. Cell 42: 5-6.
- Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 113:237-251.
- Sanger, F., S. Nicklem, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Smith, G. E., and M. D. Summers. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. Anal. Biochem. 109:123-129.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
- Stoltz, D. B., and S. B. Vinson. 1979. Viruses and parasitism in insects. Adv. Virus Res. 24:125-171.
- Theilmann, D. A., and M. D. Summers. 1986. Molecular analysis of Campoletis sonorensis virus DNA in the lepidopteran host, Heliothis virescens. J. Gen. Virol. 67:1961–1969.
- Vinson, S. B. 1972. Effect of the parasitoid, Campoletis sonorensis, on the growth of its host, Heliothis virescens. J. Insect Physiol. 18:1509–1514.
- Vinson, S. B., K. M. Edson, and D. B. Stoltz. 1979. Effect of a virus associated with the reproductive system of the parasitoid wasp, *Campoletis sonorensis*, on host weight gain. J. Invert. Pathol. 34:133-137.
- Wilson, A. C., S. S. Carlson, and T. White. 1977. Biochemical evolution. Annu. Rev. Biochem. 46:573–639.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.