From: Rickettsiae and Rickettsial Diseases (1991), pages 162-169
J. Kasar and D. Raoult (eds.). Publishing House of The Slovak Academy of Sciences, Bratislava

DNA SEQUENCE DIFFERENTIATION IN NORTH AMERICAN SPOTTED FEVER GROUP SPECIES OF RICKETTSIA

P. A. FUERST and K. F. POETTER

Dept. of Molecular Genetics, The Ohio State University, Columbus, Ohio, 43210 USA

Methods of molecular biology have recently been applied to the study of the phylogenetic relationships between different strains of tick borne Rickettsia which have been classified into the spotted fever group (SFG) /3,4/. These studies used restriction fragment pattern polymorphisms (also called RFLP patterns) of bacterial DNA to differentiate and classify species and to estimate the levels of interstrain variability within an SFG species. The members of the SFG were found to have very low levels of genetic differentiation, whether measured intraspecifically or when different species are compared. The degree of genetic variability detected in our studies of six of the SFG species is, in fact, lower than that found normally within single species of enteric bacteria, or in other free living bacteria such as Bacillus /2,5/.

It is possible that estimates of genetic variation based on restriction fragment comparisons are biased, due either to the low G+C ratio of the genome for Rickettsia (approximately 30%), or because of the relatively small size of the rickettsial genome. Consequently, we are interested in determining the level of interspecific genetic differentiation directly from DNA sequence comparisons. Since our RFLP comparisons sampled a random set of genetic segments of the rickettsial chromosome, we felt that it was important that any direct studies of gene sequences likewise reflect random portions of the genome. We present here data from three gene regions which were sampled in the RFLP studies of Poetter et al./4/. The results show that RFLP comparisons of the North American members of the SFG provided valid estimates of the genetic differentiation of the genome. The data also provide us with several insights about the patterns of gene change in the evolution of this group.

MATERIALS AND METHODS

The strains chosen as the source of genetic clones used for DNA sequencing were: R. rickettsii = Bitterroot strain, collected in 1945; R. rhipicephali = strain 3-7^{8}-6,

collected in 1973; R. parkeri = strain MAC-20, collected in 1948; R. montana = 83-441, collected from a Dermacentor variabilis tick in Ohio in 1983; R. bellii = strain 369-C collected in 1966 from a D. variabilis tick. Studies of RFLP variability within species of Rickettsia in our laboratory indicate that these strains are representative of their source species.

Interspecific variability of the nucleotide sequence was determined by dideoxy DNA sequencing of two portions of the rickettsial genome. An EcoRI fragment of the genome of R. rhipicephali was cloned in the plasmid pRHC7. The insert is 9,575 bases, and cross hybridizes with all members of the SFG and with R. bellii. This region of the rickettsial genome will be referred to as C7. HindIII fragments from the genomes of R. parkeri, R. rickettsii, R. montana, and R. bellii were ligated into the HindIII site of the plasmid pBlueskript (SK+), and the plasmids were transformed into the E. coli strain JM83. Recombinant clones for the four species which were homologous to a 2200 bp HindIII fragment of clone C7 from R. rhipicephali were identified by colony hybridizations. The homologs vary in size, with the inserts from R. parkeri and R. montana being 2150 base, R. rickettsii and R. rhipicephali being 2000 bases and R. bellii being 3450 bases in length. DNA sequences were determined for the first 350 bases at the 5' and 3' ends of each clone. Because of the difference in size compared to SFG clones, the 3' region of the C7 homolog from R. bellii was not compared with the other species.

In addition, we have compared the intraspecific variation of a second cloned segment, designated pRBE1, from two strains of R. bellii, the 369-C strain mentioned above and a second strain, 85-1251, isolated from a D. variabilis tick in Ohio in 1985. Three HindIII fragments of the insert of 369-C in plasmid pRBE1 and the genetic homologs from strain 85-1251 were subcloned, and sequenced. The total cloned fragment from strain 369-C is 8.0 kilobases in size; we have sequenced and compared 1180 bases of the three subclones from each strain.

RESULTS AND DISCUSSION

The sequences for the C7 genetic region in the four SFG species and R. bellii are given in FIGURE 1. A summary of sequence differentiation is given in TABLE 1. The primary question addressed with this data concerns possible misinterpretations of

genetic changes in Rickettsia which might occur when restriction fragment comparisons are used to estimate genetic divergence. The values of genetic differences in TABLE 1 are in good agreement with those obtained by Poetter et al./4/ using restriction fragments. Genetic differences between SFG species were all in the range 1-4% using either method. Intraspecific sequence variability in R. bellii is given in FIGURE 2. The estimate of sequence difference between the two strains (0.005) is very consistent with the value obtained using RFLP's (0.007). The comparisons indicate that estimates of genetic variability obtained from restriction fragments are not biased in studies involving Rickettsia. Since the RFLP data of Poetter et al. /4/ represent samples of over 2,000 nucleotides each, they should provide reliable estimates of differentiation when used for phylogenetic studies of these bacteria.

TABLE 1 - GENETIC DIFFERENTIATION FOR THE C7 REGION
Figures above diagonal = number of nucleotides in comparison.
Figures below diagonal = proportion of sequence differences

	R.ri.	R.p.	R.rh.	R.m.	R.b.
R. rickettsii	-	475	407	392	349
R. parkeri	0.042	-	419	395	232
R. rhipicephali	0.039	0.041	-	216	273
R. montana	0.008	0.025	0.014	-	148
R. bellii	0.329	0.353	0.315	0.250	-

weighted distance of SFG species to R. bellii = 0.319
sequence difference between strains of R. bellii = 0.0051

The sequence data summarized in TABLE 1 support previous contentions /3/ that the SFG group of Rickettsia shows extremely low levels of interspecific genetic differentiation. Studies of different strains of E. coli found levels of sequence differences in the range 2-5% /2,5/. This is similar, or even greater than the differentiation which is seen throughout the Spotted Fever Group (TABLE 1 and /4/). In contrast, the substantial level of genetic differentiation observed in comparisons of sequences from R. bellii with those of the SFG indicate that mutational changes occur

in a "normal" manner in Rickettsia and that low variability between SFG species is not due to abnormally low mutational input. Rather, lack of differentiation is more likely to be a result of very recent divergence of the SFG species from one another.

The sequences also allow us to investigate whether the mutational changes in the genome in Rickettsia occur in a biased manner. A summary of the mutational changes observed in this data set is given in TABLE 2. It can be seen that transitions (mutations involving either a pyrimidine/pyrimidine interchange or a purine/purine interchange) exceed transversions (changes from a purine to a pyrimidine, or the reverse). This excess of transitions is in disagreement with the prediction that general mutational patterns should yield transversions approximately twice as frequently as transitions. The mutational patterns for Rickettsia are very similar to those observed in vertebrate mitochondrial DNA where transitions also exceed transversions in closely related sequence comparisons /1,6/.

TABLE 2 - PATTERNS OF NUCLEOTIDE CHANGES BETWEEN SEQUENCES

	type of mutational change		
	transition	transversion	IN/DEL
R. bellii vs. SFG species	48.5	42.5	13
different SFG species	28.3	16.7	2
strains of R. bellii	2	4	0

The RFLP study of Poetter et al. /4/ found few examples of insertion/deletion (IN/DEL) changes in the genomes of Rickettsia. Very small IN/DEL events are difficult to identify in restriction fragment data, however. Examination of the sequence data from FIGURES 1 and 2, summarized in TABLE 2 supports our previous conclusion that small IN/DEL events are rare in comparisons of closely related sequences. Only two IN/DEL events have been identified in the SFG data, and none occur between the longer sequences of two strains of R. bellii (FIGURE 2). With increasing sequence differentiation, i.e. in the comparison of R. bellii with the SFG species, IN/DEL events become more common, although still representing only about 10% of mutational changes. The data for the SFG species show fewer IN/DEL events than were

observed for similar sequences in E. coli /5/, but we reserve final judgement because of the small sample of IN/DEL events observed. However, the data of Poetter et al. /4/, which indicates that large IN/DEL events are also rare in Rickettsia, differs substantially from observations in E. coli /5/, where large events are more common.

The sequences in FIGURE 1 provide our first quantitative estimates of the genetic differences between R. bellii and the members of the SFG. Other studies in our laboratory, especially studies of gene sequences from the 16S rRNA genes of R. bellii, R. montana and R. rickettsii have already suggested that R. bellii is intermediate between the typhus and spotted fever groups, although it is clearly closer to the tick borne SFG. The data presented here, combined with the lack of common restriction fragments between R. bellii and the SFG /4/, show clearly that R. bellii diverged from other tick borne rickettsiae before the divergence of SFG species from each other.

CONCLUSIONS

DNA sequences from the spotted fever group rickettsiae have been compared to estimate genetic divergence. Estimates of genetic differentiation between species are similar to estimates reported earlier based on the comparison of DNA restriction fragment patterns. Genetic differences between species of the spotted fever group of Rickettsia are extremely small, and probably reflect a very recent divergence within the tick borne rickettsias.

REFERENCES

1. Brown, W.M. /1985/: in: Molecular Evolutionary Genetics, R.J. MacIntyre (ed.), Plenum Pr., New York, pp. 95-130.

2. Dubose, R.F., Dykhuisen, D.E. and Hartl, D.L. /1988/: Proc. Nat. Acad. Sci. USA 85: 7036-7040.

3. Fuerst, P.A., Poetter, K.P., Pretzman, C. and Perlman, P.S. /1990/: Ann. New York Acad. Sci. 590: 430-438.

4. Poetter, K.P., Pretzman, C., Ralph, D., Perlman, P.S. and Fuerst, P.A. /1990/: J. Bacteriology (submitted).

5. Stoltzfus, A., Leslie, J.F. and Milkman, R. /1988/: Genetics 120: 345-358.

6. Wilson, A.C., et al. /1985/: Biol. J. Linnean Soc. 26: 375-400.

Figure 1. Sequence comparisons of C7 homologs of North American species. Alignments were done by the program GENEPRO (Riverside Scientific). A period (.) designates a site which differs only in R. bellii. An asterisk (*) designates a site which is polymorphic among SFG species. The sequences in the top of the figure represent aligned sites beginning at the 5' end of the homologous cloned segment, while the sequences at the bottom represent the sequence at the 3' end of clones. Because of a difference in the size of the clones, the 3' end of the C7 homolog from R. bellii was not compared with the other species. Ri = R. rickettsii, Rh = R. rhipicephali, P = R. parkeri, M = R. montana, and B = R. bellii.

FIGURE 1. SEQUENCE COMPARISONS OF C7 HOMOLOGS IN NORTH AMERICAN SFG SPECIES.

A. SEGMENT I: 354 base region at 5' end of subclone:

```
      . .. . .   *          .       .   . .*.   *       .                . . . .                 . ..  * . .. .    .*   .     .
Ri    TAAGACGACG AGCAGGTTTA CCGTATTCTC CGGTGACAGA TAGTTTTAAG TCTGCATCTT CGCTAATTTC ATGTTCATAA GTTAAGCCGG CATAAAAAGC ATCCCTAATA TTCTGATTAA
P     TAAGACGACG CGCAGGTTTA CCGTATTCTC CGGTGGCAGA TAATTTTAAG TCTGCATCTT CGCTAATTTC ATGTTCATAA GTTAAGCCGG CAGAAAAAGC ATCCCTAATA TTCTGATTAA
Rh                                                                                 GTTCATAA GTTAAGCCGG CAGAAAAAGC ATCCTTAATA TTCTGATTAA
M     TAAGACGACG AGCAGGTTTA CCGTATTCTC CGGTGACAGA TAGTTTTAAG TCTGCATCTT CGCTAATTTC ATGTTCATAA GTTAAGCCGG CATAAAAAGC ATCCCTAATA TTCTGATTAA
B     CAGCAAGGCG AGCAGGTTTA CCATATTCAC CAGTAACGGA TAATTTTAAA TCTGCATCTT CACCAAGTGC ATGTTCATAA GTTAAACCAC CAGATATTGA ATCTGTAACA TTTTGATTAA

        .  ..       .. .    ... *. ..      . *  ....   . .. .   .* . *.   ..... * ... *  ... *  ....           .    . .. .   *
Ri    TAGTCATGGT TTCGTTATTC TCTAATTTAA CCGTCTTTAT TCCGGTTTTA GAAA-CACTA ACTCC-GTT- TCTGCCTGAG CTTTCAAGCG TTA---AATT ATTT-ATAT- -CTCTATTAT
P     TAGTCATGGT TTCGTTATTC TCTAATTTAA CCGTCTTTAT TCCAGTTTTA GAAA-CACTA ACCCC-GTC- TCTGCCTGAA CTTTTAAGCG ATA---AATT ATTT-ATAT- -CTCTATTAC
Rh    TAGTCATGGT TTCGTTATTC TCTAATTTGA CCGTCTTTAT TCCGGTTTTA GAAA-CACTA ACTCC-GTT- TCTGCCTGAG CTTTCAAGCG -TA---AATT ATTT-ATAT- -CTCTATTAC
M     TAGTCATGGT TTCGTTATTC TCTAATTT
B     TAGTAATAAT TTCGTCGTTT TCTACAATGT TAGTCTTTAT GCCAGT-GCG GACATTAGTA ATACCAGTAA TC-----GAA TCATCAA--- ATACGACATT ATTTCATATT TTTC-ATTAC

      *** *  . * *   .   .            . ..         * .            ... .  .   ..*           *.          .      . . ...*
Ri    TAGCAGTATC GGCAGTATCG GGAGTATAAG AAACTCCGGC TTGGATC-TT TCATTTTAGG CGTGAAAAAA TTTAATCTTC TAGCTCTCTC AGCTTTATCA TTGATTTGAA CAAA
P
Rh    CTCCTGTATT AGCAGTATCG GGAGTATAAG AAACTCCGGC TTGGAGC-TT TCATTTTAGG CGTGAAAAAA TTCAATCTTC TAGCTTTCTC AGCTTTATCA TTGATTTGAT CAAA
M
B     CACCTGTGTT GGCAGAATCA GGAGTATAAG ATATACCGGC TTGGAACCTT TCATTTTAGG --GGATAGAA GCTAATCTTC TAGCTGCCTC AGCTTTAACA TTAAATGCGT CAAA
```

B: SEGMENT II: 368 base segment at 3' end of subclone:

```
      **                      *                      *                                                         *   *
Ri    GTTCAATTAA GTTAGCCATT TCCGGCTCGC CTTCAATTAA CAAAACTCGC ATTATTTTTC TCCATAAAAT TCAATTTTTA TATTGTGCTA TAATTTAATT ACAGGTAATT ATTGTTAACT
P     CATCAATTAA GTTAGCCATT TCCGGCTCAC CTTCAATTAA CAAAACTCTC ATTATTTTTC TCCATAAAAT TCAATTTTTA TATTGTGCTA TAATTTAATT ACAGTTAACT ATTGTTAACT
Rh                                                                                                                     ATT ATTGTTAACT
M     GTTCAATTAA GTTAGCCATT TCCGGCTCGC CTTCAATTAA CAAAACTCTC ATTATTTTTC TCCATAAAAT TCAATTTTTA TATTGTGCTA TAATTTAATT ACAGTTAATT ATTGTTAACT

                             *                      *                 ***                                     *
Ri    AAATTATAAT TTTTTATCTA GTGCTTTATA CATTCATTAA TAATATAAGT AAAGAATATT TAATC---T TGATAAAATTT TTACTATTAA GCTCTATCTG TGGTTTAAAA ACCATGGCGT
P     AAATTATAAT TTTTTATCTA GCGCTTTATA CATTCATTAA TAACATAAGT AAAGAATATT TAATCAAAT TGATAAAATTT TTACTATTAA GCTCTATCTG TGATTTAAAA ACCATGGCGT
RH    AAATTATAAT TTTTTATCTA GTGCTTTATA CATTCATTAA TAACATAAGT AAAGAATATT TAATCAAAT TGATAAAATTT TTACTATTAA GCTCTATCTG TGATTTAAAA ACCATGGCGT
M     AAATTATAAT TTTTTATCTA GTGCTTTATA CATTCATTAA TAACATAAGT AAAGAATATT TAATCAAAT TGATAAAATTT TTACTATTAA GCTCTATCTG TGGTTTAAAA ACCATGGCGT

            *                   *                  *  *  *                     *                 *   *
Ri    TTTAAAA
P     TTTAAAACGA AATTTATTAA ATAATTAAAT GATTTTCTTA GGTTTTGATG TTCCGACGAA ATTTATTAAA TAATTAAATG ATTTTCTTAG GTTTTGATGT TCCGACATTT TCTTATGATA
Rh    TTTAAAACAA AATTTATTAA ATAATTAAAT GACTTTCTTA GGTTTTGATG CTCTGACAAA ATTTATTAAA TAATTAAATG ACTTTCTTAG GTTTTGATGC TCTGACATTT TCTTATGATA
M     TTTAAAA

P     TAAGGCAC
Rh    TAAGGCAC
```

Figure 2. Comparison of R. bellii sequences from strains 369-C and 85-1251. Three homologous subcloned segments are designated I,II,and III in this figure. T7 and T3 designate the sequencing primers used to obtain each sequence. Polymorphic sites are given in lowercase and underlined. Sequence listed is that from strain 369-C.

Overall total: 1186 sites, 6 differences; nucleotide difference = 0.0051

I. Clone E1/H3 subclone #1;

A. T7 primer: 192 sites, 0 differences

GGAGATTTTG CCTTTATTTT AGGCGTAATA ACGATTATTT TTTATTGTCA TTCAGCAAAT
TACGAGGATG TATTTTTGCT TGCTCCAAAA TTAGCTAATA CAAAAATATT GCTAGCTGAT
TTTGAAATTT CTATCCTTGA TATTGCCTGC TTATTGTTAT TCATCGGTTT GTATGGGTAA
ATCGGCACAG AT

B. T3 primer: 192 sites, 0 differences

GTCTAGTTTT GCCGTGGAAT ACCAACATAA TAATTTTCAT TGAGTAAATA GCGGTAAGTA
TTGCAGCTAA TATACCAAAT ATGAACATAA ATGACCCGCT GCTATAAGCC GCTTCCAAGA
TCGAATCTTT TGAATAAAAC CCTGCAAGCG GATAGATCCC TATTAACGCA AGCGAACCGA
TTAAAAAATT TC

II. Clone E1/H3 subclone #2;

A. T7 primer: 217 sites, 1 difference

TTTATTTTTg TTAAGCAGCA TCATTAACAC CGCCAGCTAG CTTATTTGAA ATTTCTCTAG
AACGCTCAAC TATCTTCTTA ATATCAATTT TAGGCAGAAC CAACATCAAC GCCATAATCT
TTAGCGTGTT TTATATACTC AAACACTTCA GCAGATTTAA GCAAGGATTT CGTCGGAATA
CATCCCAATT AAGGCACACT CCACCCAAAT GCTCTTT

B. T3 primer: 247 sites, 4 differences

ATAGCTACGC CTTATATTTa AAAACACAAA ATTATcATTG GAATATTGAA GGAGCTGAAT
TTAGAAGCCT GCATTTATTA TTCGAAGGGC AATATGAAgA TTTGgCAGAA AGTCTAGATG
AGCTAGCTGA ACGAATAAGA ACTTTAGATG CTAAAGTTCC GACATTATCA AATTTGATAA
AACTAGCATC TATAAGTGAG CCAAACCCTA ATGCTTCTGC AAATGAAATG CTAAAAAGCC
TTGTAAA

III. E1/H3 subclone #3;

A. T7 primer: 150 sites, 0 differences

ATACCACGAT TTAGGTGAAT AAATTTATTA CTATTGTAGT TATTGTTGCA ACTAATATTA
CTAGGCATCT ATATTATTAT TCCACTATCC CCGCCTCAAC AATACAATAT ATCGCCGTCT
TAAATTGTTG TTATATAGTT TGTTTCCGTG

B. T3 primer: 188 sites, 1 difference

CTGACCAACT GCTCTACTTG CATCATCTAG CTTCTCATTT ACGTCTTTAA CTGTTTTTTT
TATTCTGTTT TTCATATAGA CCTCAATTTA AgGTTCTATA ACTAGAGTAA CACTTGAATC
TTTTATAATC AACTAGAAAA ATTTAACTAA CAAATTACTC TACTGACAAA ATGAAATTAC
ATTTAAAA

Polymorphic differences in Ohio isolate, strain 85-1251:
subclone 2;T7, site 10 = A;
subclone 2;T3, site 20 = C, site 36 = A, site 99 = T,
site 105 = A;
subclone 3;T3, site 92 = A