

9. T. Kunik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1871 (2001).
10. J. M. Young, L. D. Kuykendall, E. Martinez-Romero, A. Kerr, H. Sawada, *Int. J. Syst. Evol. Microbiol.* **51**, 89 (2001).
11. E. Martinez, R. Palacios, F. Sanchez, *J. Bacteriol.* **169**, 2828 (1987).
12. F. Galibert *et al.*, *Science* **293**, 668 (2001).
13. T. Kaneko *et al.*, *DNA Res.* **7**, 331 (2000).
14. The complete annotated sequence, detailed methods, and all supplementary data, which will be periodically updated, are available at our Web site (www.agrobacterium.org). Supplemental data are also available at Science Online (www.sciencemag.org/cgi/content/full/294/5550/2317/DC1). The sequence has been deposited at GenBank (accession numbers: circular chromosome, AE008688; linear chromosome, AE008689; pAtC58, AE008687; pTiC58, AE008690).
15. At a late phase of this project, we became aware of an independent effort by Goodner *et al.* (52) to sequence the C58 strain of *A. tumefaciens*. Comparison of the two independently finished sequences indicates excellent overall agreement. There are two small insertions in the other sequence, one in the circular chromosome and one in pAtC58, relative to our sequence. Single nucleotide discrepancies are far below current accuracy standards for finished genomic sequence (99.99%). Further work will determine which of the observed discrepancies reflect strain differences and which reflect sequencing errors.
16. R. H. Hamilton, M. Z. Fall, *Experientia* **27**, 229 (1971).
17. A. Allardet-Servent, S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan, M. Ramuz, *J. Bacteriol.* **175**, 7869 (1993).
18. J. Zhu *et al.*, *J. Bacteriol.* **182**, 3885 (2000).
19. V. S. Kalogeraki, S. C. Winans, *J. Bacteriol.* **180**, 5660 (1998).
20. R. L. Tatusov *et al.*, *Nucleic Acids Res.* **29**, 22 (2001).
21. J. A. Eisen, J. F. Heidelberg, O. White, S. L. Salzberg, *Genome Biol.* **1**, Research 0011.1 (2000).
22. C. M. Fraser *et al.*, *Nature* **390**, 580 (1997).
23. J. N. Volff, J. Altenbuchner, *FEMS Microbiol. Lett.* **186**, 143 (2000).
24. K. Bao, S. N. Cohen, *Genes Dev.* **15**, 1518 (2001).
25. M. A. Ramirez-Romero, N. Soberon, A. Perez-Oseguera, J. Tellez-Sosa, M. A. Cevallos, *J. Bacteriol.* **182**, 3117 (2000).
26. L. Chen, unpublished data.
27. S. K. Farrand, in *The Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 199–233.
28. I. T. Paulsen, L. Nguyen, M. K. Sliwinski, R. Rabus, M. H. Saier Jr., *J. Mol. Biol.* **301**, 75 (2000).
29. J. A. Forward, M. C. Behrendt, N. R. Wyborn, R. Cross, D. J. Kelly, *J. Bacteriol.* **179**, 5482 (1997).
30. R. Sanchez-Fernandez, T. G. Davies, J. O. Coleman, P. A. Rea, *J. Biol. Chem.* **276**, 30231 (2001).
31. C. K. Stover *et al.*, *Nature* **406**, 959 (2000).
32. D. Missiakas, S. Raina, *Mol. Microbiol.* **28**, 1059 (1998).
33. G. T. Robertson, R. M. Roop Jr., *Mol. Microbiol.* **34**, 690 (1999).
34. A. G. Matthysse, G. W. Kijne, in *The Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 235–249.
35. K. J. Fullner, J. C. Lara, E. W. Nester, *Science* **273**, 1107 (1996).
36. P. L. Li, I. Hwang, H. Miyagi, H. True, S. K. Farrand, *J. Bacteriol.* **181**, 5033 (1999).
37. J. R. Harper, T. J. Silhavy, in *Principles of Bacterial Pathogenesis*, E. A. Groisman, Ed. (Academic Press, San Diego, CA, 2001), pp. 43–74.
38. J. H. Weiner *et al.*, *Cell* **93**, 93 (1998).
39. K. LeVier, R. W. Phillips, V. K. Grippe, R. M. Roop II, G. C. Walker, *Science* **287**, 2492 (2000).
40. M. J. Pallen, B. W. Wren, *Mol. Microbiol.* **26**, 209 (1997).
41. A. Ichige, G. C. Walker, *J. Bacteriol.* **179**, 209 (1997).
42. S. A. Coleman, M. F. Minnick, *Infect. Immun.* **69**, 4373 (2001).
43. M. Carsiotis, B. A. Stocker, D. L. Weinstein, A. D. O'Brien, *Infect. Immun.* **57**, 3276 (1989).
44. Y. Dessaux, A. Petit, S. K. Farrand, P. J. Murphy, in *The*

- Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 173–197.
45. V. S. Kalogeraki, J. Zhu, A. Eberhard, E. L. Madsen, S. C. Winans, *Mol. Microbiol.* **34**, 512 (1999).
46. K. J. Miller, J. M. Wood, *Annu. Rev. Microbiol.* **50**, 101 (1996).
47. Z. Q. Luo, S. K. Farrand, *J. Bacteriol.* **181**, 618 (1999).
48. B. Minambres, E. R. Olivera, R. A. Jensen, J. M. Luengo, *J. Biol. Chem.* **275**, 39529 (2000).
49. E. Jumas-Bilak, S. Michaux-Charachon, G. Bourg, M. Ramuz, A. Allardet-Servent, *J. Bacteriol.* **180**, 2749 (1998).
50. H. Ochman, N. A. Moran, *Science* **292**, 1096 (2001).

51. J. A. Downie, J. P. Young, *Nature* **412**, 597 (2001).
52. B. Goodner *et al.*, *Science* **294**, 2323 (2001).
53. We thank P. Green for useful discussions, R. Gibson for providing the genome_plot program, and J. Staley for critical review of the manuscript. Supported by NSF grant 0092815 (M.V.O.), NIH Public Health Service grant GM32618 (E.W.N.), NIH doctoral fellowship GM19642 (D.W.W.), Sao Paulo State Research Foundation (FAPESP) sabbatical fellowship 1999/11876-5 (J.C.S.), a FUNDECT-MS grant (N.F.A.), and the E. I. du Pont de Nemours Company.

3 October 2001; accepted 9 November 2001

Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58

Brad Goodner,^{1,2} Gregory Hinkle,³ Stacie Gattung,⁴ Nancy Miller,⁴ Mary Blanchard,⁴ Barbara Quorollo,³ Barry S. Goldman,^{3,4} Yongwei Cao,³ Manor Askenazi,³ Conrad Halling,³ Lori Mullin,³ Kathryn Houmiel,⁴ Jeffrey Gordon,³ Mark Vaudin,⁴ Oleg Iartchouk,³ Andrew Epp,³ Fang Liu,³ Clifford Wollam,⁴ Mike Allinger,⁴ Dahlia Doughty,² Charlaïne Scott,² Courtney Lappas,² Brian Markelz,² Casey Flanagan,² Chris Crowell,² Jordan Gurson,² Caroline Lomo,² Carolyn Sear,² Graham Strub,² Chris Cielo,² Steven Slater^{3*}

Agrobacterium tumefaciens is a plant pathogen capable of transferring a defined segment of DNA to a host plant, generating a gall tumor. Replacing the transferred tumor-inducing genes with exogenous DNA allows the introduction of any desired gene into the plant. Thus, *A. tumefaciens* has been critical for the development of modern plant genetics and agricultural biotechnology. Here we describe the genome of *A. tumefaciens* strain C58, which has an unusual structure consisting of one circular and one linear chromosome. We discuss genome architecture and evolution and additional genes potentially involved in virulence and metabolic parasitism of host plants.

Agrobacterium tumefaciens is a plant pathogen with the unique ability to transfer a defined segment of DNA to eukaryotes, where it integrates into the eukaryotic genome. This ability to transfer and integrate DNA is used for random mutagenesis and has been adapted into a powerful tool for production of transgenic plants, including soybean, maize, and cotton (1, 2). *A. tumefaciens* was identified early in the 20th century as the causal agent of crown gall disease in plants (3). Pathogenesis is initiated when *Agrobacte-*

rium detects small molecules released by actively growing cells in a plant wound. These molecules induce a series of virulence (*vir*) genes whose encoded products export the single-stranded transferred DNA (T-DNA) to the plant cell, where it integrates into the genome at an essentially random location. Once integrated, T-DNA gene expression alters plant hormone levels, leading to cell proliferation typical of a gall tumor. The T-DNA also encodes enzymes for the synthesis of opines, a class of nutrient molecules used almost exclusively by *A. tumefaciens* (4–7).

A. tumefaciens strains fall into three biovars, which differ in their host range, metabolic characteristics, relationships with other genera in the family Rhizobiaceae, and potentially their chromosome structure (4–13). The taxonomy of the Rhizobiaceae family is not without controversy, but we expect that

¹Department of Biology, Hiram College, Hiram, OH 44234, USA. ²Department of Biology, University of Richmond, Richmond, VA 23173, USA. ³Cereon Genomics, LLC, 45 Sidney Street, Cambridge, MA 02139, USA. ⁴Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, MO 63167, USA.

*To whom correspondence should be addressed. E-mail: steven.c.slater@cereon.com

phylogenetic analysis that includes genome sequences should help resolve this issue (12, 13). Until that time, we believe that the biovar concept is the most valuable for understanding the diversity within the genus *Agrobacterium*. We analyzed *A. tumefaciens* strain C58, a biovar 1 strain with an unusual genome structure known only in this genus: one circular chromosome and one linear chromosome. All pathogenic strains also harbor a tumor-inducing (Ti) plasmid that encodes the T-DNA and virulence genes, and they are classified by the types of opines induced and metabolized (4–7). Some strains also harbor additional plasmids, but previous studies focused heavily on Ti plasmids, including the recently completed sequences of a nopaline-agrocinopine-type plasmid, pTiSAKURA, and a consensus plasmid based on several closely related octopine-type plasmids (4–6, 14–16). *A. tumefaciens* C58, isolated from a cherry tree (*Prunus*) tumor, carries two plasmids: pAtC58 and the nopaline-agrocinopine-type pTiC58 (17). This strain has been intensively studied and is the parent of many strains used for the genetic transformation of plants (2, 4–6).

Agrobacterium is an α -proteobacterium within the Rhizobiaceae family, an agriculturally important group that includes the nitrogen-fixing symbiotic partners of legumes (18). Six complete genomes are now available within the α -proteobacteria: *Sinorhizobium meliloti* and *Mesorhizobium loti* (both nitrogen-fixing symbionts), *Caulobacter crescentus*, two *Rickettsia* species, and *A. tumefaciens* (19–24). The mammalian pathogens *Brucella* and *Bartonella* are also related, so comparative genomics may reveal commonalities and differences between pathogens and symbionts (25, 26).

Overall genome structure. The *A. tumefaciens* C58 genome consists of a circular chromosome, a linear chromosome, and two plasmids: the tumor-inducing plasmid pTiC58 and a second plasmid, pAtC58 (8, 10). The genome was sequenced and assem-

bled with standard methods (27). The sequence of all four DNA molecules is available on GenBank, along with chromosomal diagrams and analysis tools (28). The essential features of the four molecules are summarized in Table 1 and in the accompanying Web materials (27). The two chromosomes contain all of the genes for stable RNAs and housekeeping proteins involved in essential cellular functions and prototrophic growth.

The circular chromosome contains a putative origin of replication (*Cori*) similar to the known *Cori* of *Caulobacter crescentus* (29). The linear chromosome, on the other hand, has a plasmid-type replication system of the same type found on pTiC58 and pAtC58. This system, encoded by the *repABC* genes, expresses a pair of segregation proteins (RepA and RepB) and an origin-binding replication initiation protein (RepC) (30). Thus, we hypothesize that the linear chromosome is evolutionarily derived from a plasmid. The plasmid origin of an “extra”

chromosome had been predicted for multichromosome genomes of the α -proteobacteria and has been found in more distantly related organisms such as *Vibrio cholerae* (19, 31–33).

Gene density is very similar between the two chromosomes. However, genes involved in most essential processes are significantly overrepresented on the circular chromosome (34) (Fig. 1). This asymmetry is consistent with direct descent of the circular chromosome from the primordial α -proteobacterial genome, with a minority of essential genes moving to the linear chromosome (10). Consistent with lateral transfer between chromosomes over a long period, the overall dinucleotide signatures of the two chromosomes are essentially identical but are significantly different from those of the two plasmids (27, 35). The dinucleotide signatures of the two plasmids are quite similar to each other and to related plasmids from other members of the Rhizobiaceae family (27).

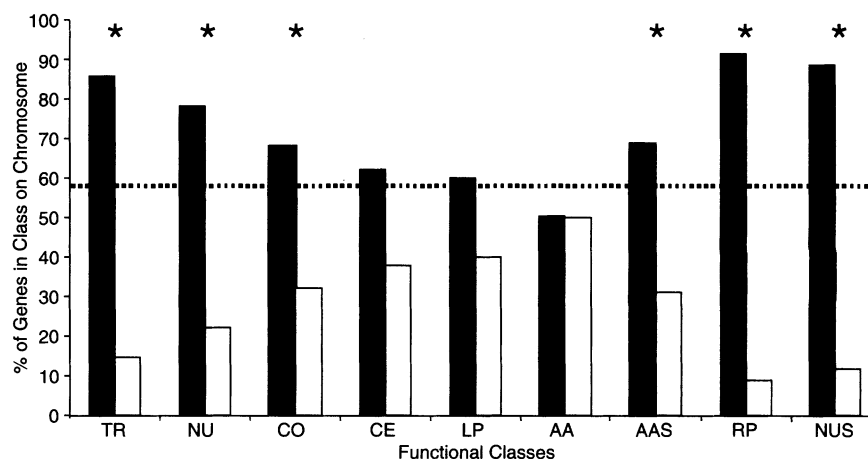


Fig. 1. Distribution of genes from different functional classes on the circular chromosome (solid bars) versus the linear chromosome (open bars). Clusters of orthologous groups (COG) (34) classes are as follows: TR, translation; NU, nucleotide metabolism; CO, coenzyme metabolism; CE, cell envelope; LP, lipid metabolism; and AA, amino acid metabolism. Other classes are AAS, amino acid synthesis; RP, ribosomal proteins; and NUS, nucleotide synthesis. Asterisks indicate significant differences in gene distribution (goodness of fit test). The dotted line represents the null hypothesis of gene distribution based on chromosome size for the circular chromosome.

Table 1. General features of the *A. tumefaciens* genome.

	Circular chromosome	Linear chromosome	pAtC58	pTiC58	Total
Size (bp)	2,841,581	~2,075,600	542,869	214,233	~5,674,300
G+C content	59.4%	59.3%	57.3%	56.7%	
Stable RNAs					
rRNAs	2 <i>rrn</i> operons	2 <i>rrn</i> operons	–	–	
tRNAs	40 species	13 species	–	–	
Coding regions*	2722 (89.1%)	1833 (91.2%)	547 (85.6%)	197 (86.2%)	5299 (89.8%)
Known function†‡	60 (2.2%)	29 (1.5%)	51 (9.3%)	88 (44.7%)	228 (4.3%)
Probable function†§	1708 (62.7%)	1268 (69.2%)	272 (49.7%)	79 (40.1%)	3327 (62.8%)
Possible function†	124 (4.6%)	78 (4.3%)	25 (4.6%)	7 (3.6%)	234 (4.4%)
Conserved unknown†¶	152 (5.6%)	79 (4.3%)	29 (5.3%)	3 (1.5%)	263 (5.0%)
Novel†#	678 (24.9%)	379 (20.7%)	170 (31.1%)	20 (10.1%)	1247 (23.5%)

*Number of predicted open reading frames for that DNA molecule (percentage of DNA involved in coding). †Number of predicted open reading frames for that DNA molecule (percentage of total). ‡Known *A. tumefaciens* genes with demonstrated function. §Overall similarity ($P < 1 \times 10^{-15}$) to genes with demonstrated or probable function from other organisms. ||Genes where limited similarity or short motifs suggest a function. ¶Strong similarity ($P < 1 \times 10^{-15}$) to genes with unknown function from other organisms. #Novel genes.

More than 6000 base pairs of near-perfect sequence identity extend across the two ribosomal RNA (rRNA) gene clusters on each of the two chromosomes. The chromosomes also share some shorter regions of greater than 90% sequence identity with pAtC58 (27). Transcription of all rRNA gene clusters is oriented away from the DNA replication origins, with those on the linear chromosome in the same orientation. A number of house-keeping genes are located between the linear chromosome's rRNA operons, and one might expect frequent recombination resulting in lethal events.

The telomeres of the linear chromosome are covalently closed. The linear chromosome is a covalently closed linear molecule, apparently possessing hairpin loops at the telomeres (Fig. 2). Our sequence comes within several hundred bases of the telomeres, and additional sequence is presented by Wood *et al.* (24). However, the sequence of the putative hairpin loop is not yet available. The proximal regions of both telomeres are similar in overall architecture but very different in sequence. These regions contain portions of several insertion sequence (IS) elements with intervening DNA of additional repeated and unique sequence. They are rich in potential secondary structure and contain numerous short sequence repeats.

Precedence for covalently closed telomeres is established in *Escherichia coli* phage N15 and spirochaetes such as *Borrelia burgdorferi* (36, 37). DNA replication proceeds through the hairpin, and telomere duplicates are separated to permit chromosome segregation. Resolution of duplicated telomeres is best characterized in the phage N15

system, where protelomerase separates the telomere copies, then rejoins the ends of individual molecules to create hairpin loops (36). No ortholog of N15 protelomerase was identified in the *A. tumefaciens* C58 genome. However, several putative transposases are encoded near the telomeres, and these could separate daughter telomeres in a reaction analogous to transposon excision.

DNA replication and the cell cycle.

DNA replication is synchronized with the *A. tumefaciens* cell cycle (11, 38, 39), a feat requiring coordination of four DNA molecules and two different classes of replication origin: *Cori* and *repABC*. The precise signal that initiates DNA replication is not yet clear, although in the related α -proteobacterium *Caulobacter*, many proteins are subject to cell cycle-synchronized transcriptional control and proteolysis (40). Because the initiation signal must be interpretable by both types of replication origin, it is unlikely to be transduced by a single origin-specific binding protein.

Processive DNA replication is performed by DNA polymerase III (Pol III), and *A. tumefaciens* carries four paralogs of the *dnaE* gene encoding the Pol III α (polymerase) subunit (41). These *dnaE* genes fall into two distinct sequence families, designated as categories A and B (27). The category A gene of the circular chromosome is conserved in all sequenced α -proteobacteria and probably encodes the primary replication enzyme (19–23, 27, 41). Each of the *A. tumefaciens* *repABC* replicons (linear chromosome, pTiC58, and pAtC58) encodes a category B *dnaE* gene within an operon containing two conserved genes of unknown function. The

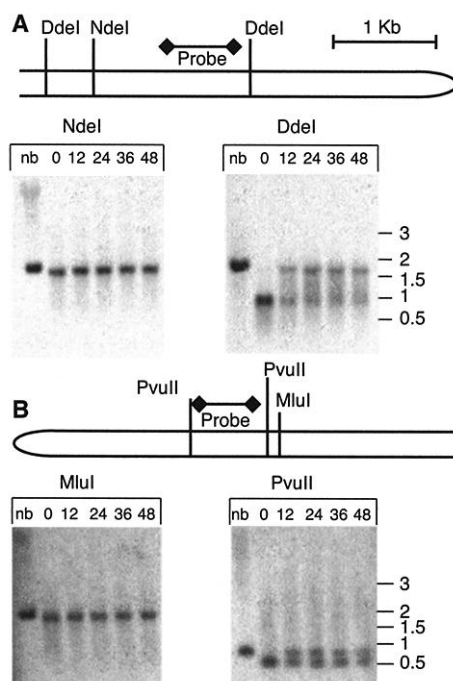
operon is present in all fully sequenced α -proteobacteria except the *Rickettsia* species and may encode a novel DNA polymerase complex (19–23).

Syntenic analysis of the *A. tumefaciens*, *S. meliloti*, and *M. loti* genomes.

Although *Agrobacterium* and *Sinorhizobium* are in different phylogenetic clades (13), there is an excellent syntenic relationship between the *A. tumefaciens* circular chromosome and the chromosome of *S. meliloti*. This further supports the idea that the circular chromosome is the original chromosome involved during the evolutionary radiation of the α -proteobacteria, with the subsequent evolution of the linear chromosome taking place from a plasmid (19, 42) (Fig. 3A). Many of the *S. meliloti* orthologs missing from the circular chromosome are present on the linear chromosome, as are a number of gene duplications. The region of *S. meliloti* retaining the most synteny with the *A. tumefaciens* linear chromosome contains about 300 genes, with orthologs arranged in short syntenic groups across the linear chromosome (19, 42) (Fig. 3A). These genes retain the same gene order relative to *S. meliloti*, despite their broad dispersal. If portions of the linear chromosome arose via an excision event from the primordial chromosome, the excision may have originated in this region, with subsequent insertions moving particular sections apart. Many other *S. meliloti* orthologs are also present on the linear chromosome, but their seemingly random location suggests many independent transfers. IS elements are relatively rare on the linear chromosome (with the exception of the telomeres) and cannot explain the highly distributed nature of orthologous genes. The synteny between the circular chromosome of *A. tumefaciens* and the chromosome of *M. loti* is much less pronounced and deserves further analysis in terms of the evolutionary history of the family Rhizobiaceae (20) (Fig. 3D).

All four *A. tumefaciens* replicons contain some genes similar to those on the *S. meliloti* pSymA and pSymB plasmids (19, 27, 31, 43) (Fig. 3, B and C). Most of these shared genes encode members of large orthologous groups, such as ABC transporters. However, there are exceptions, including the *fixNOQP* operon found on the circular chromosome in *A. tumefaciens* (44). In *S. meliloti*, these genes lie on pSymA and encode a special cytochrome oxidase required for the microaerophilic growth found in symbiotic nodules. The role of this cytochrome *cbb₃*-type oxidase in the biology of *A. tumefaciens* awaits further study. As with comparisons of the *A. tumefaciens* linear chromosome versus the *S. meliloti* chromosome, orthologous genes from the pSym plasmids are spread across both *A. tumefaciens* chromosomes in many short regions of similarity.

Fig. 2. The linear chromosome telomeres are covalently closed, as shown by Southern blot hybridization using probes near the right (A) and left (B) telomeres. Chromosomal DNA was digested with specified restriction endonucleases in 100 mM NaCl, 50 mM tris-Cl, 10 mM MgCl₂, and 1 mM dithiothreitol, and a portion of each reaction mix was boiled for 10 min, then allowed to cool slowly. After electrophoresis and blot transfer, specific probes identified DNA fragments containing intact telomeres (Nde I and Mlu I digests) or lacking telomeres (Dde I and Pvu II digests). The mobility of fragments possessing telomeres was essentially unchanged by denaturation, suggesting that telomeres contain a covalent hairpin loop. DNA lacking telomeres was denatured by boiling, creating two single-stranded molecules. Slow cooling allowed renaturation of a portion of the denatured molecules. nb, not boiled; 0, 12, 24, 36, and 48 indicate cooling time (in minutes) before freezing in a dry ice and ethanol bath. After 48 min of cooling, the temperature of the DNA samples was approximately 50°C. Numbers on the right of each figure indicate sizes (in kilobases) of double-stranded DNA molecular weight standards.



Plant transformation and tumorigenesis. Genes involved in plant transformation and tumorigenesis are located on all four genetic elements. The circular chromosome harbors the well-studied *chvAB* genes required for synthesis of the extracellular β -1,2-glucan involved in binding to plant cells; the *chvGI*, *chvE*, and *ros* genes involved in regulation of Ti plasmid *vir* genes; and the *chvD*, *chvH*, and *acvB* genes (4–6). The linear chromosome harbors the *exoC* (*pgm*) gene required for synthesis of the extracellular β -1,2-glucan and succinoglucan polysaccharides, and the cellulose synthesis (*cel*) genes involved in binding to plant cells (4–6). Our *cel* region sequence differed from the published sequence, and we have reannotated this locus (27). Plasmid pAtC58 contains the attachment (*att*) genes involved in initial specific attachment of the bacterium to plant cells, as well as a second, partial *att* locus (4–6). pAtC58 is reportedly dispensable for virulence, raising the question of whether there is a virulence requirement for *att* (14).

Ti plasmids fall into several opine groups, and the three plasmid sequences now available permit detailed analyses of their relationships (4–6, 15, 16) (Fig. 4). The order of genes on the nopaline-agrocinopine-type plasmids pTiC58 and pTi-SAKURA are virtually identical. Major exceptions include one large insertion on pTiC58 and four smaller insertions on pTi-SAKURA. In contrast, the consensus octopine-type plasmid shares only five major gene clusters with the nopaline-type plasmids (Fig. 4). Many regional differences among these plasmids can be circumstantially linked to mobile DNA elements (45). Most of the pTiC58-specific genes are involved in metabolism and transport and probably allow the bacterium to scavenge additional nutrients. However, for a few genes unique to a given Ti plasmid, there is direct or circumstantial evidence that they play a role in tumorigenesis on particular hosts (for example, *ligE* of pTi-SAKURA, and *virH1*, *virH2*, and *virJ* of octopine-type plasmids) (15, 16). Both pTiC58 and pTi-SAKURA encode a probable NUDIX hydrolase, which may degrade altered nucleotides or other toxic compounds present in the plant wound environment (46).

Secretion systems and other pathogenicity factors. *A. tumefaciens* lacks a dedicated type III secretion system for exported pathogenicity factors, but it does have the flagellar type III system, and there is evidence of chemotaxis in response to plant-released compounds (47, 48). Flagella are known to secrete pathogenicity factors in *Yersinia enterocolitica*, and flagellar mutants of *A. tumefaciens* exhibit moderately reduced virulence (49, 50). However, the *A. tumefaciens* studies do not discriminate between the

potential effects of motility and protein secretion (50).

A. tumefaciens C58 contains three type IV

secretion systems and is the primary model for understanding secretion of DNA and proteins through these transporters. Two systems

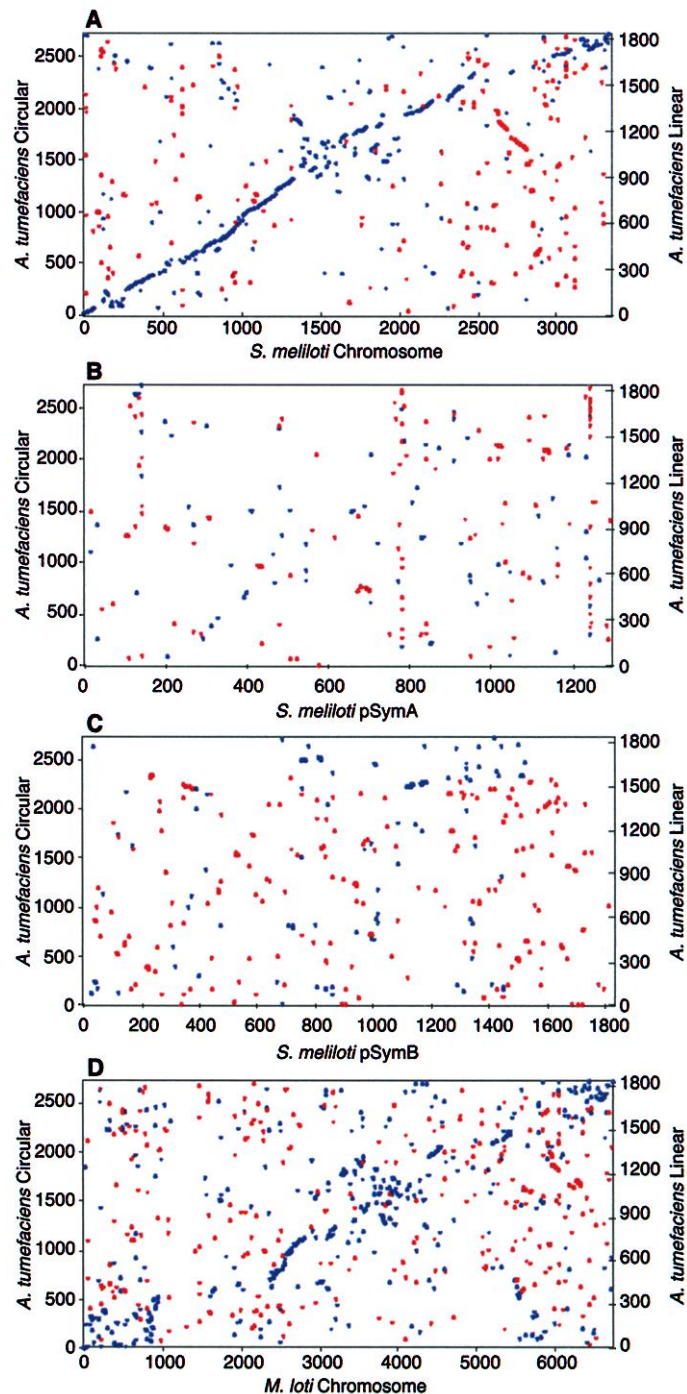


Fig. 3. Synteny among *S. meliloti*, *M. loti*, and the two *A. tumefaciens* chromosomes (19, 20). Protein pairs were generated by BLASTP analysis of predicted proteins from each genome, retaining only the best *S. meliloti* or *M. loti* match for each *A. tumefaciens* protein. Each protein pair is graphed according to the location of the corresponding gene on respective DNA molecules. Blue dots, alignment between *A. tumefaciens* circular chromosome and a designated *S. meliloti* or *M. loti* replicon. Red dots, alignment between *A. tumefaciens* linear chromosome and a designated *S. meliloti* or *M. loti* replicon. (A) Comparison of *A. tumefaciens* with *S. meliloti* chromosome. (B) Comparison of *A. tumefaciens* with *S. meliloti* plasmid pSymA. (C) Comparison of *A. tumefaciens* with *S. meliloti* plasmid pSymB. (D) Comparison of *A. tumefaciens* with *M. loti* chromosome. Synteny between *A. tumefaciens* and the *M. loti* pML plasmids was extremely low (27). To reduce background from members of orthologous groups, only pairs with a BLASTP expectation value $\leq 1 \times 10^{-60}$ are shown, but full BLASTP data are available (27).

encoded by pTiC58 have been heavily studied: the VirB system for T-DNA transfer to plant cells and the Trb/Tra system for conjugal transfer of pTiC58 (4–6). A third system is encoded by pAtC58. The utility of this system is unknown, but pAtC58 may be conjugative (51). The type IV complexes most similar to the pAtC58 system are from the related animal pathogens *Brucella* and *Bartonella* (25, 26).

Several additional loci may play a role in establishing or maintaining a pathogenic relationship with the plant. An autotransporting virulence factor family member is encoded by pAtC58. Such proteins cross the plasma membrane via the signal peptide-dependent pathway and self-insert into the outer membrane (52). Typically, a large extracellular domain is exposed, where it modifies cell adhesion or host cell functions. Other genes similar to known virulence factors include orthologs of *bacA* (macrophage survival in *Brucella abortus* and legume symbiosis in *Sinorhizobium meliloti*), putative adhesins, *icmF* (macrophage killing in *Rickettsia*), and as many as six different iron uptake systems (53, 54). Iron acquisition is always a priority for pathogens, and *A. tumefaciens* strains vary in their production of siderophores (55). Although strain C58 does not produce detectable siderophore activity in low-iron medium, the diversity of iron uptake systems may allow it to co-opt some catechol, hydroxamate, or citrate siderophores produced by other microbes or by plants (55).

Other aspects of metabolism and signaling. *A. tumefaciens* establishes a proprietary carbon and nitrogen source by genetically engineering its host to produce opines (4–7). The genes for nopaline and agropinopine utilization lie on pTiC58, but strain C58 derivatives lacking pTiC58 can take up octopine and nopaline without subsequent hydro-

lysis, and spontaneous mutants of these derivatives have been found with an octopine or mannopine/mannopinic acid utilization phenotype (56, 57). Genome analysis provides some clues about the basis of these activities. There are many putative amino acid, dipeptide, and oligopeptide ABC transport systems that may permit scavenging a variety of nitrogenous compounds. Also, the linear chromosome and pAtC58 both encode strong homologs of AgaE, the Ti plasmid-encoded version of which catalyzes the breakdown of mannopine to mannose and glutamate (58).

A. tumefaciens also exploits many native plant metabolites, such as sucrose, tannins, and cell wall polymers. Sucrose is the major form by which organic carbon is transported in most plants. Like its relative *S. meliloti*, *A. tumefaciens* has an α -glucoside utilization (*agl*) operon and an additional orphan α -glucosidase, as well as sucrose hydrolase (invertase) (59). A third route unique to *A. tumefaciens* involves the periplasmic oxidation of sucrose to 3-ketosucrose, transport of 3-ketosucrose, and cytoplasmic cleavage into fructose and 3-ketoglucose. With no homologs for comparison, we cannot positively identify the genes encoding the enzymes of this pathway, but mutants affecting the pathway are available (60). Plasmid pAtC58 encodes a protein with strong similarity to fungal tannases. This enzyme might allow the bacterium to use tannins as nutrients or to defend itself against the antimicrobial activities of many tannins (61). Finally, there are numerous genes that may allow degradation of cellulose, hemicellulose, pectin, and/or lignin. These include a β -endoglucanase encoded within the cellulose synthesis gene cluster, putative xylan esterases and xylanases, a previously identified inducible polygalacturonase (pectinase), and several enzymes for degrading monomeric and dimeric compo-

nents of plant lignin (LigE, VirH2, and the β -ketoadipate pathway) (4–6, 15, 16, 62, 63).

The *A. tumefaciens* genome encodes at least 25 different two-component regulatory systems and a wide array of other regulatory proteins, including two new bacterial phytochromes (64). Typically, bacterial phytochromes contain the sensory portion of the protein, including the tetrapyrrole chromophore-binding site, attached to a histidine kinase domain. One of the *A. tumefaciens* phytochromes has this structure, and its gene is in a putative operon with a partner response regulator. The other phytochrome is itself a response regulator. There are no data in the literature as to the effect of light on *A. tumefaciens*.

Concluding remarks. The analysis presented here and in the accompanying paper by Wood *et al.* illuminate many new avenues to further explore the biology and biotechnological utility of *A. tumefaciens* (24). These include the maintenance of a complex genome, new potential virulence mechanisms, and many additional ways in which the bacterium may parasitize a plant host. At a larger level, the availability of six α -proteobacterial genomes, with more on the way, provides a wealth of comparative data for further understanding the evolutionary history of the group, the evolution and maintenance of multichromosome genomes, and the evolution of mechanisms that interact with and exploit animal and plant hosts (19–26).

References and Notes

1. J. F. Topping, W. Wei, M. C. Clarke, P. Muskett, K. Lindsey, *Methods Mol. Biol.* **49**, 63 (1995).
2. C. A. Newell, *Mol. Biotechnol.* **16**, 53 (2000).
3. E. F. Smith, C. O. Townsend, *Science* **25**, 671 (1907).
4. A. N. Binns, M. F. Thomashow, *Annu. Rev. Microbiol.* **42**, 575 (1988).
5. Y. Dessaux, A. Petit, S. K. Farrand, P. J. Murphy, in *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria*, H. P. Spink, A. Kondorosi, P. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 173–197.
6. J. Zupan, T. R. Muth, O. Draper, P. Zambryski, *Plant J.* **23**, 11 (2000).
7. M.-D. Chilton *et al.*, *Cell* **11**, 263 (1977).
8. A. Allardet-Servent, S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan, M. Ramuz, *J. Bacteriol.* **175**, 7869 (1993).
9. E. Jumas-Bilak, S. Michaux-Charachon, G. Bourg, M. Ramuz, A. Allardet-Servent, *J. Bacteriol.* **180**, 2749 (1998).
10. B. W. Goodner *et al.*, *J. Bacteriol.* **181**, 5160 (1999).
11. K. Suzuki, K. Iwata, K. Yoshida, *DNA Res.* **31**, 141 (2001).
12. A. Willems, M. D. Collins, *Int. J. Syst. Bacteriol.* **43**, 305 (1993).
13. J. M. Young, L. D. Kuykendall, E. Martinez-Romero, A. Kerr, H. Sawada, *Int. J. Syst. Evol. Microbiol.* **51**, 89 (2001).
14. C. Rosenberg, T. Huguet, *Mol. Gen. Genet.* **196**, 533 (1984).
15. K. Suzuki *et al.*, *Gene* **242**, 331 (2000).
16. J. Zhu *et al.*, *J. Bacteriol.* **182**, 3885 (2000).
17. R. H. Hamilton, M. Z. Fall, *Experientia* **27**, 229 (1971).
18. V. Oke, S. R. Long, *Curr. Opin. Microbiol.* **2**, 641 (1999).
19. F. Galibert *et al.*, *Science* **293**, 668 (2001).
20. T. Kaneko *et al.*, *DNA Res.* **7**, 381 (2000).

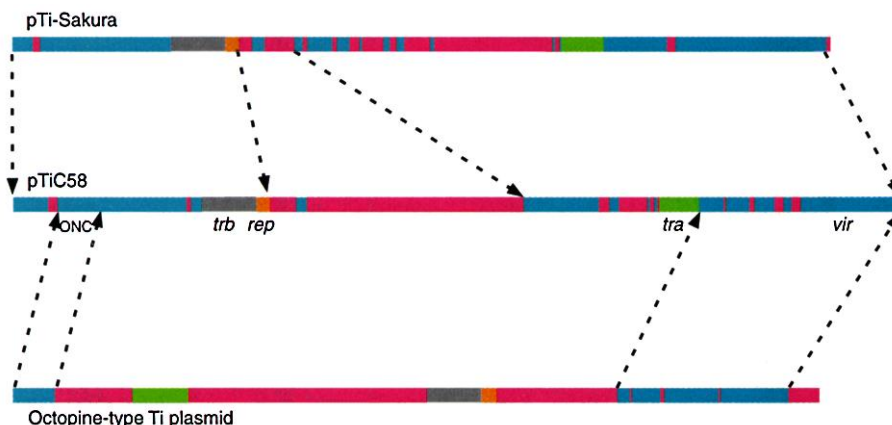


Fig. 4. Schematic linear alignment, beginning at the left T-DNA border, of pTiC58 with the two other sequenced Ti plasmids (15, 16). Blue represents areas of homologous genes in the same or similar location. Gray (*trb*), orange (*rep*), and green (*tra*) represent areas of homologous genes that are in different locations on various plasmids. Red represents unique regions on each plasmid. The positions of the T-DNA oncogenes (*ONC*), *trb* operon, *rep* operon, *tra* operons, and *vir* operons are shown for pTiC58. Arrows delineate regions of identical gene order among the plasmids.

21. W. C. Nierman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4136 (2001).
22. S. G. Andersson *et al.*, *Nature* **396**, 133 (1998).
23. H. Ogata *et al.*, *Science* **293**, 2093 (2001).
24. At a late phase of this project, we became aware of an independent effort to sequence the *A. tumefaciens* C58 genome. The work of these authors is published jointly in this issue [D. Wood *et al.*, *Science* **294**, 2317 (2001)]. The two independently derived sequences are in excellent overall agreement, with single nucleotide discrepancies far below current accuracy standards (99.99%) for finished sequence. Our sequence contains two small insertions relative to that of Wood *et al.*; one each in the circular chromosome and pAtC58. Further work will determine which of the observed discrepancies reflect strain differences and which reflect errors.
25. M. L. Boschioli, V. Foulongne, D. O'Callaghan, *Curr. Opin. Microbiol.* **4**, 58 (2001).
26. K. L. Karem, C. D. Paddock, R. L. Regnery, *Microbes Infect.* **2**, 1193 (2000).
27. Supplementary Web material is available on Science Online at www.sciencemag.org/cgi/content/full/294/5550/2323/DC1.
28. DNA sequence and protein predictions are available at the National Center for Biotechnology Information's Entrez Genomes Server (www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html). The GenBank accession numbers are NC_003062 (circular chromosome), NC_003063 (linear chromosome), NC_003064 (pAtC58) and NC_003065 (pTiC58).
29. A. K. Brassinga, R. Siam, G. T. Marczyński, *J. Bacteriol.* **183**, 1824 (2001).
30. P. L. Li, S. K. Farrand, *J. Bacteriol.* **182**, 179 (2000).
31. T. M. Finan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9889 (2001).
32. E. Moreno, *FEMS Microbiol. Rev.* **22**, 255 (1998).
33. J. F. Heidelberg *et al.*, *Nature* **406**, 477 (2000).
34. R. L. Tatusov, *Nucleic Acids Res.* **29**, 22 (2001).
35. S. Karlin, *Trends Microbiol.* **9**, 335 (2001).
36. J. Deneke, G. Ziegelin, R. Lurz, E. Lanka, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7721 (2000).
37. S. Casjens *et al.*, *Mol. Microbiol.* **26**, 581 (1997).
38. L. S. Kahng, L. Shapiro, *J. Bacteriol.* **183**, 3065 (2001). Kahng and Shapiro analyzed the appearance of hemimethylated DNA, upon passage of a replication fork, at the *att* locus in *A. tumefaciens*. At the time of their study, *att* was believed to reside on the circular chromosome, but we correctly locate it on pAtC58. Thus, at least one *repABC* origin, and probably all three, is coordinated with the cell cycle.
39. K. Skarstad (personal communication) has shown by flow cytometry that all DNA replication is restricted to a relatively short time span within the cell cycle of *A. tumefaciens*.
40. I. J. Domian, A. Reisenauer, L. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6648 (1999).
41. Z. Kelman, M. O'Donnell, *Annu. Rev. Biochem.* **64**, 171 (1995).
42. D. Capela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9877 (2001).
43. M. J. Barnett *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9883 (2001).
44. A. Schluter, S. Ruberg, M. Kramer, S. Weidner, U. B. Priefer, *Mol. Gen. Genet.* **247**, 206 (1995).
45. L. Otten *et al.*, *Mol. Plant Microbe Interact.* **5**, 279 (1992).
46. M. J. Bessman, D. N. Frick, S. F. O'Handley, *J. Biol. Chem.* **271**, 25059 (1996).
47. G. V. Plano, J. B. Day, F. Ferracci, *Mol. Microbiol.* **40**, 284 (2001).
48. C. H. Shaw, *Bioessays* **13**, 25 (1991).
49. G. M. Young, D. H. Schmiel, V. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6456 (1999).
50. E. M. Lai, O. Chesnokova, L. M. Banta, C. I. Kado, *J. Bacteriol.* **182**, 3705 (2000).
51. M. Van Montagu, J. Schell, in *Plasmids of Medical, Environmental & Commercial Importance*, K. M. Timmis, A. Puhler, Eds. (Elsevier, Amsterdam, 1979), pp. 71–95.
52. I. R. Henderson, F. Navarro-Garcia, J. P. Nataro, *Trends Microbiol.* **6**, 370 (1998).
53. K. LeVier, R. W. Phillips, V. K. Grippe, R. M. Roop 2nd, G. C. Walker, *Science* **287**, 2492 (2000).
54. M. Purcell, H. A. Shuman, *Infect. Immun.* **66**, 2245 (1998).
55. R. Penyalver, P. Oger, M. M. Lopez, S. K. Farrand, *Appl. Environ. Microbiol.* **67**, 654 (2001).
56. M. Krishnan, J. W. Burgner, W. S. Chilton, S. B. Gelvin, *J. Bacteriol.* **173**, 903 (1991).
57. G. LaPointe, C. S. Nautiyal, W. S. Chilton, S. K. Farrand, P. Dion, *J. Bacteriol.* **174**, 2631 (1992).
58. S. M. Lyi, S. Jafri, S. C. Winans, *Mol. Microbiol.* **31**, 339 (1999).
59. L. B. Willis, G. C. Walker, *J. Bacteriol.* **181**, 4176 (1999).
60. P. L. Schuerman, J. S. Liu, H. Mou, A. M. Dandekar, *Appl. Microbiol. Biotechnol.* **47**, 560 (1997).
61. T. K. Bhat, B. Singh, O. P. Sharma, *Biodegradation* **9**, 43 (1998).
62. S. Subramaniyan, P. Prema, *FEMS Microbiol. Lett.* **183**, 1 (2000).
63. D. Parke, *J. Bacteriol.* **177**, 3808 (1995).
64. R. D. Vierstra, S. J. Davis, *Semin. Cell Dev. Biol.* **11**, 511 (2000).
65. Supported by funds from Cereon Genomics; Monsanto Company; the University of Richmond; and the National Institutes of Health, National Institute of General Medicine (grant R15 GM61690-01 to B.W.G.). We thank the University of Richmond students of the 1999 and 2000 genetics course and the 2000 and 2001 microbiology course for assistance; S. Salzberg, C. Frasier, and The Institute for Genome Research for providing GLIMMER; and L. S. Kahng, J. Elhai, R. Last, J. Hirschberg, A. Matthyse, S. Winans, S. Farrand, and the manuscript's reviewers for helpful discussions, well-reasoned criticisms, and corrections to our original annotations.

3 October 2001; accepted 19 November 2001

Earthquake Recurrence and Rupture Dynamics of Himalayan Frontal Thrust, India

Senthil Kumar,^{1*} Steven G. Wesnousky,¹ Thomas K. Rockwell,² Daniel Ragona,² Vikram C. Thakur,³ Gordon G. Seitz⁴

The Black Mango fault is a structural discontinuity that transforms motion between two segments of the active Himalayan Frontal Thrust (HFT) in north-western India. The Black Mango fault displays evidence of two large surface rupture earthquakes during the past 650 years, subsequent to 1294 A.D. and 1423 A.D., and possibly another rupture at about 260 A.D. Displacement during the last two earthquakes was at minimum 4.6 meters and 2.4 to 4.0 meters, respectively, and possibly larger for the 260 A.D. event. Abandoned terraces of the adjacent Markanda River record uplift due to slip on the underlying HFT of 4.8 ± 0.9 millimeters per year or greater since the mid-Holocene. The uplift rate is equivalent to rates of fault slip and crustal shortening of $9.6^{+7.0}_{-3.5}$ millimeters per year and $8.4^{+7.3}_{-3.6}$ millimeters per year, respectively, when it is assumed that the HFT dips $30^\circ \pm 10^\circ$.

The ongoing collision of India into Eurasia has produced four major thrust earthquakes along the ~2500-km length of the Himalayan front during the past ~100 years (1), yet none of the events reportedly produced coseismic surface ruptures (2, 3) (Fig. 1A). Here, we report paleoseismological evidence of the size and timing of surface-rupturing thrust earthquakes along the HFT as well as a quantitative bound on the rate of fault slip on the HFT, from a site located about 80 km northwest of Dehra Dun.

Regional setting. The collision of India with Eurasia since the Eocene has accommodated 2000 to 3000 km of convergence (4). The boundary between the Indian and Eurasian plates forms an arc that extends ~2500 km across the continent (Fig. 1A). Plate mo-

tion models and recent Global Positioning System (GPS) measurements indicate that convergence between the Indian and Eurasian plates is between about 40 and 50 mm/year (5, 6). Between about 10 and 20 mm/year of the total 40 to 50 mm/year of convergence is taken up by thrust motion along the Himalayan arc, with the remainder taken up farther to the north by a combination of thrusting, crustal extension, and strike-slip motion within the Eurasian Plate (3, 7–9). Three major south-verging thrust faults strike the length of the Himalayan arc (Fig. 1). The northernmost of the three thrusts is the Main Central Thrust (MCT), which emerges along the southern edge of the High Himalaya and has not been observed to break Quaternary deposits and, hence, is generally considered inactive (10). The Main Boundary Thrust (MBT) marks the southern edge of the Lesser Himalaya, is expressed in bedrock along the arc, and is locally observed to displace Quaternary deposits (10, 11). The southernmost thrust is the HFT, which is now considered to be the most active of the three and delineates the northern limit of the exposed Indian Plate.

¹Center for Neotectonic Studies, University of Nevada, Reno, NV 89557, USA. ²Department of Geological Sciences, San Diego State University, San Diego, CA 92182, USA. ³Wadia Institute of Himalayan Geology, 33, G. M. Singh Road, Dehra Dun 248001, UP, India. ⁴Lawrence Livermore National Laboratory, Livermore, CA 94551, USA.

*To whom correspondence should be addressed. E-mail: senthil@seismo.unr.edu