Virulence shift in a sexual clade of Type X *Toxoplasma* infecting Southern Sea Otters

KENNARD, A[1], MILLER, MA[2], KHAN, A[1*], QUINONES, M[3], MILLER, N[4], SUNDAR, N[1], JAMES, ER[5], ROOS, DR[4], CONRAD, PA[6] AND GRIGG, ME[1,5].

[1]Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, MD, USA. [2]California Department of Fish and Game, Santa Cruz, CA, USA. [3]Bioinformatics and Computational Biosciences Branch, NIAID, National Institutes of Health, Bethesda, MD, USA. [4]Department of Biology and Penn Genome Frontiers Institute, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [5]Department of Medicine, University of British Columbia, Vancouver, BC, Canada. [6]School of Veterinary Medicine, University of California, Davis, CA, USA.

*Current Address: Agricultural Research Service, USDA, Beltsville, MD, USA.

Abstract

How virulent protozoal pathogens capable of causing overt disease are maintained in nature is an important paradigm of eukaryotic pathogenesis. Here we used population genetics and molecular methods to study the evolution and emergence of a marine invasion of new genetic variants of *Toxoplasma gondii,* referred collectively as Type X (HG12). 53 *Toxoplasma* isolates were obtained from mustelids that stranded between 1998-2004 with toxoplasmosis (ranging from chronic infection to fatal encephalitis). Over 74% of the sea otters collected throughout their geographic range were infected with Type X as determined by multi-locus PCR-DNA sequencing. Depending on the locus investigated, Type X strains possessed one of three allelic types that had independently assorted across the strains examined; either genetically distinct alleles, referred to as "γ" or "δ", or a Type II allele. Phylogenetic incongruence among locus-specific trees, genome-wide CGH array and WGS analyses confirmed that Type X is a sexual clade of natural recombinants that resemble F1 progeny from a genetic cross between Type II and a mosaic of two distinct "γ" or "δ" ancestries. A single Type X genotype (19/53; 36%) had expanded in sea otters largely as subclinical chronic infections, but it was highly pathogenic to mice (LD_{100}= 1 parasite). To determine whether murine virulence genes could be mapped within this naturally occurring population, we performed a genome scan and identified four QTLs with LOD scores greater than 4.0. Targeted disruption of ROP33, the strongest candidate from among 16 genes within the highest QTL on Chromosome VIIa established ROP33 as a murine virulence locus. The ability of this highly pathogenic clone to expand and cause the majority of sea otter infections supports a virulence shift model whereby generalist pathogens like *Toxoplasma* utilize their sexual cycles to produce new strains with an expanded biological potential. Such a trait enables pathogens to extend their host range or be naturally selected within their vast intermediate host range to maximize their transmission. Our work thus establishes a rationale for how virulent strains can be maintained cryptically in nature across a pathogen's broad host range, and act as reservoirs for epidemic disease.

Importance

Waterborne outbreaks of protozoal parasites are increasingly causing fatal disease in a wide range of animals, including humans. Population expansion of felids near marine estuarine environments has led to increased exposure of marine wildlife to highly infectious *Toxoplasma gondii* oocysts shed in the feces of cats that are dispersed by storm events. In North America *Toxoplasma* is thought to possess a highly clonal population structure

dominated by 4 clonal lineages (I, II, III, and X). Population genetic analysis of 53 *Toxoplasma* isolates collected longitudinally from mustelids infected with *Toxoplasma* that stranded between 1998-2004 identified a majority of otters (74%) to be infected with Type X *Toxoplasma*, and that Type X is not a clonal lineage, but rather a recombinant clade of strains consistent with a recent genetic cross that produced at least 12 distinct haplotypes. Importantly, one Type X haplotype expanded in 36% of otters across their geographic range and caused relatively benign infections, however it was highly pathogenic to mice. A genome scan was performed to identify a new virulence locus, a secreted serine threonine kinase (ROP33), that is pathogenic in mice, but not sea otters. Our data support a virulence shift model whereby generalist pathogens like *Toxoplasma* utilize their sexual cycles to produce virulent strains that are maintained cryptically in nature, according to their differential capacity to cause disease within the pathogen's broad intermediate host range. This type of "zoonotic selection" has important public health implications. Strains capable of causing fatal infections can persist in nature by circulating as chronic infections in intermediate host species that act as reservoirs for epidemic disease.

Introduction

How successful microbes maintain reservoirs of virulent strains in nature is understudied and is an important paradigm of infectious diseases. Whereas many studies document how microbial pathogens acquire or admix genetic material to influence their pathogenicity, relatively less is known how emergent strains are selectively maintained in nature. Specifically, it is well established that viruses utilize reassortment to rapidly produce admixture lines that possess a range of altered biological potential, including virulence [1]. Bacteria transfer mobile plasmid elements that encode pathogenicity islands via conjugation to confer an altered virulence potential on those strains that receive them [2-4]. Among both fungi and protozoa, genetic hybridization by sexual recombination is a potent way to produce genomic variation or new admixture lines that possess increased virulence and are capable of causing disease outbreaks or altering host range [5-7]. What is less clear are the mechanisms at play allowing highly prevalent pathogens to maintain virulent strains cryptically in nature. We sought to test whether generalist pathogens, such as the protozoan parasite *Toxoplasma gondii*, recovered from natural infections in one animal host, leverages its broad range of intermediate hosts to selectively partition parasite genotypic diversity and virulence potential.

Toxoplasma gondii is a highly successful and prevalent protozoan pathogen that infects a wide range of wildlife, livestock and 20-80% of humans worldwide [8-10]. Its success in nature is largely attributed to its highly flexible life cycle. It is propagated asexually by carnivory among essentially all warm-blooded vertebrates (via orally infectious tissue cysts) or sexually within its definitive felid host by self-mating (when a single parasite undergoes fertilization and sexually clones itself) or by out-crossing (producing as many as 10^8 genetic hybrids that are transmissible as environmentally stable, highly infectious oocysts). Genetic hybridization has been previously shown to produce genotypes that possess a wide spectrum of biological potential, including altered virulence or a capacity to expand as successful epidemic clones [7, 11, 12]. Importantly, parasite clones possess a broad spectrum of disease states that are highly dependent on both parasite genetics and the animal species infected [8, 13-15]. In rodents, low dose inoculums of Type I *Toxoplasma* strains are uniformly lethal to laboratory mice [16] whereas Type II strains are considerably

less virulent and routinely establish chronic, transmissible infections [14]. In contrast, wild-derived CIM mice that encode resistant alleles of immunity-related GTPases (IRGs) confer resistance to virulent Type I strains by inactivating parasite-specific secreted kinases (ROP18/ROP5) that control mouse susceptibility to infection [17, 18]. The end result of this host genetic, parasite genetic interplay determines the potential for a disease-producing virulent clone to expand or be maintained cryptically within a strain-specific, or host species-specific, manner, but this has not so far been demonstrated to occur in nature and has only been inferred by laboratory studies. Likewise, laboratory rats are resistant to Type I infections and fail to transmit parasites, but again this is entirely rat species-dependent [19]. The molecular basis for this resistance is an allele-dependent activation of the NLRP1 inflammasome resulting in macrophage pyroptosis and the inhibition of parasite growth in Lewis (LEW) and Sprague Dawley (SD) rats, but not in Brown Norway (BN) or Fischer CDF rats, which go on to establish chronic, transmissible infections of mouse-virulent Type I strains [20-22]. While these laboratory studies suggest the possibility that virulent strains can be maintained cryptically across the pathogen's host range [23], the phenomenon has not been systematically investigated nor demonstrated to occur in a natural setting.

Between 1998 and 2004, *Toxoplasma gondii* emerged as a significant disease in California sea otters (*Enhydra lutris nereis*), a federally listed threatened species [24, 25]. Full necropsies performed on each stranded sea otter identified a wide spectrum of disease sequelae among *Toxoplasma*-infected otters, ranging from chronic asymptomatic infection to fatal meningoencephalitis at time of death [25, 26]. Importantly, the majority of sea otters were found to be infected by a single outbreak clone, referred to as Type X (or HG12) [24]. Based on multi-locus PCR-DNA sequencing (MLST) using a limited set of genotyping markers, Type X was defined as the fourth clonal lineage in the US [27] and it is highly prevalent in sylvatic niches as chronic, subclinical or mild infections in 47% of US wildlife [28, 29]. The range of disease sequelae detected in sea otters, however, was not consistent with infection by a single clone. Other possibilities proposed to explain variation in disease susceptibility have included co-infection [30], exposure to environmental pollutants, or toxins that cause immunosuppression [31]. More recently, WGS performed on *Toxoplasma* isolates has called into question whether MLST analyses are sufficiently resolved to predict clonotypes because they fail to capture extant genetic heterogeneity [15]. Indeed, a previous study concluded that Type X was comprised of two genotypes [32], but these two genotypes differed by only a few SNPs, so it was not clear whether strain variation was likewise playing a contributing role in sea otter disease in this novel host-parasite interaction.

To test for this and to determine the true extent of genetic diversity among naturally circulating Type X isolates, we sequenced a wide selection of linked and unlinked markers across the nuclear and organellar genomes for 53 isolates collected longitudinally from infected sea otters across a seven year period. Based on these results, we selected 16 isolates for WGS. Analysis of our data support a model whereby Type X exists as a recombinant clade of strains that resemble F1 progeny from a natural cross between a Type II strain and a novel genotype that has not been identified previously in nature. We show that one sea otter genotype is widely distributed and has expanded to cause the majority of infections, which were largely subclinical and consistent with a strain that has been naturally selected in sea otters. This genotype, however, was highly pathogenic to laboratory mice and a population level, forward genetic approach using WGS mapped a mouse virulence gene to a novel parasite-specific secreted kinase ROP33, also referred to as WNG3 [33]. Our natural

population dataset argues that *Toxoplasma* leverages its broad range of intermediate hosts to partition genetic diversity. Further, that intermediate hosts play a central role in the natural selection, expansion and maintenance of cryptically virulent strains in a host species-specific manner and that these infected hosts act as reservoirs for pathogenic or epidemic disease in another host species.

Results

Genetic characterization of *Toxoplasma gondii* strains isolated from sea otters

In the late 1990s, an increasing number of sea otters stranded along the California coast with protozoal encephalitis due to *Toxoplasma gondii* [25, 29, 32, 34]. An initial genetic analysis performed on a subset of 35 *T. gondii* isolates established that 21 were a new genetic type, referred to as Type X (also known as Haplogroup 12, or HG12), but this conclusion was limited to sequence typing at a single locus, GRA6 [25]. We developed a multi-locus (MLST) PCR-DNA sequence-based typing scheme using an additional 4 unlinked markers (BSR4, BAG1, SAG3, and ROP1) that possess a wide-range of phylogenetic strength, including markers under neutral (BSR4), diversifying (GRA6, SAG3, ROP1), and purifying (BAG1) selection to determine the molecular genotypes of 53 *T. gondii* isolates collected longitudinally over a seven-year period between 1998-2004 from fresh, beachcast California sea otters. Isolates were obtained throughout the California sea otter range (Supplemental Figure 1). Antibody titers against *Toxoplasma* (when available), stranding location, otter sex, and co-infection status with other protozoal agents was available for 52 of the sea otters (Supplemental Figure 1).

Point-source outbreak of Type II strains infecting sea otters

To determine the *T. gondii* genotypes among the sea otter isolates, the 5 locus MLST was applied. Typing at just the GRA6 marker identified three sequence types, either a Type II allele in 16 (30%) isolates, or two non-archetypal alleles, previously identified as X or A [25, 32], in 25 (47%) and 12 (23%) isolates, respectively. Expanding the genetic analysis across five markers showed that 2 of 16 isolates with a Type II allele at GRA6 were predicted to be non Type II, recombinant strains, because they possessed non-archetypal alleles at ROP1 (3675) or BSR4 and BAG1 (3671) (Figure 1A). Of the remaining 14 otters infected with Type II strains, 12 (86%) were identified within a restricted geographic range, a 40 mile stretch off the California coast line containing Moss Landing and Monterey Bay, which is consistent with a point source outbreak of a *Toxoplasma* strain that is found most frequently infecting human and domestic livestock in the USA. The other two Type II infected otters were recovered from Pismo Beach in Central California. Further, the majority of Type II infected otters (9/14; 64%) were collected during the first half of the study (Supplemental Figure 1). Males were predominantly infected with this genotype, and it was the only genotype that had a male-specific infection sex bias. [Fisher's two-sided exact test II vs. all; p=0.008]

Type X is a recombinant clade of strains

Of the remaining 39 (74%) *Toxoplasma* isolates, collectively referred to as Type X or A by the allele present at GRA6, a maximum of only three alleles were identified at the additional sequenced loci, either a canonical Type II allele or one of two genetically distinct alleles,

186 referred to as "γ" or "δ", that appeared to segregate independently across the isolates (Figure
187 1A). In total 8 distinct haplotypes designated A-H were resolved based on their differential
188 inheritance of the limited alleles, consistent with recombination between a Type II strain and
189 a mosaic of two distinct ancestries as the most plausible explanation for the genetic
190 relationship among the non-Type II strains. The Type X haplotype designated A was
191 dominant (19/39; 49%). It was widely distributed across the entire geographic range of the
192 Southern sea otter (Supplemental Figure 1), and it was isolated in every year except 2000.
193 Importantly, this genotype had expanded during the time frame of the study period, and it
194 was responsible for the majority of infections by 2003 (Figure 1B).
195
196 **Acute virulence in mice is dependent on Type X haplotype**
197 Mouse virulence is highly dependent on *Toxoplasma* genotype [7, 16, 35]. To further
198 characterize the otter isolates, 18 strains, including two Type II and at least one isolate
199 representing each of the 8 distinct Type X haplotypes, were tested for their mouse virulence
200 phenotype. Isolates were selected based largely on their ability to expand *in vitro* in human
201 foreskin fibroblasts. Virulence was assayed in female CD-1 outbred mice using a low-dose,
202 50 tachyzoite, intra-peritoneal injection model. At this inoculum, mouse avirulent strains
203 (*e.g.* Type II) establish chronic, transmissible infections whereas mouse virulent strains (*e.g.*
204 Type I) die acutely within 10-14 days.
205 Among the otter isolates, three mouse virulence phenotypes were identified: virulent
206 (red; all mice died acutely), intermediate virulent (blue; some mice survived acute disease),
207 and avirulent (green; all mice survived acute infection and produced chronic disease). The
208 two otter isolates that possessed a Type II MLST were avirulent (Figure 1C; green) and
209 phenocopied the infection kinetics of two well studied Type II lines Me49 and 76K. Among
210 the Type X haplotypes, E and G were also avirulent. Three otter isolates from haplotype A,
211 the genotype that had expanded to cause the majority of sea otter infections, were highly
212 pathogenic to mice; all infected mice died acutely (Figure 1C; red). Haplotypes D and H were
213 likewise highly virulent in mice. The three haplotype F otter isolates all possessed an
214 intermediate virulence phenotype (Figure 1C; blue). Of particular note, the 5 isolates from
215 haplotypes B and C had a mix of two mouse virulence phenotypes, they were either avirulent
216 (green) or they were intermediate virulent (blue) (Figure 1C). Because the Type X isolates
217 displayed a range of virulence phenotypes, the data collectively support a model whereby
218 Type X represents a clade of recombinants from a genetic admixture of a mouse avirulent
219 Type II strain with at least one other parent that is mouse virulent, rather than expansion of
220 a single clonal lineage. Furthermore, not all isolates within each subgroup displayed
221 equivalent virulence kinetics, in particular isolates from the B and C haplotypes were either
222 avirulent or possessed intermediate virulence, suggesting that the 5 marker MLST was
223 insufficiently resolved to fully capture total genetic and phenotypic diversity among the
224 isolates recovered from infected Southern sea otters.
225 To explore this possibility, the 18 isolates assayed through mice plus 3 additional
226 isolates, one each from haplotypes A, C and D, were further genotyped at 18 markers (17
227 nuclear-encoded single copy gene loci plus one microsatellite marker) representing an
228 expanded set of linked and unlinked loci across the genome. Additionally, 2 markers, one
229 each from the organellar genomes of the apicoplast and mitochondria, were included to
230 determine the ancestry of each maternally inherited genome. This represents a significant
231 increase in the number of markers that were previously used to conclude that Type X (HG12)

232 is a clonal lineage [10, 27, 36]. Altogether, the 20 marker MLST method provided 15,404 bp
233 of sequence information to determine the genetic relationship among *Toxoplasma* strains
234 isolated from Southern sea otters (Supplemental Table 1).
235
236 **Type X is comprised of 12 distinct haplotypes by an expanded MLST analysis**
237 To establish the number of haplotypes among the 21 isolates, sequence data for the
238 17 nuclear-encoded single copy gene loci was concatenated and analysed using eBURST
239 (Figure 1D). eBURST resolves strains into clonal complexes (CC) that are in linkage
240 disequilibrium. In this case, coded MLST data was used to identify strains that differ at one
241 locus (or share 16 out of 17 markers) and isolates that differed at the same genetic locus
242 were grouped together and connected by a line within each clonal complex. As expected,
243 eBURST clustered the two Type II isolates (2987, 3131) into a single clonal complex with
244 Me49 (CC1). Type X subgroup G isolate 3675, originally identified by its allele at the
245 microsatellite marker ROP1, was included within CC1 because it differed from Type II strains
246 at only SAG4. For the remaining 18 Type X isolates, 15 clustered into 3 clonal complexes
247 (designated CC2-CC4), and 3 unique haplotypes were resolved. The eBURST analysis failed
248 to support a clonal lineage designation for Type X. Further, the increased resolution provided
249 by the 17 markers subdivided the 8 distinct haplotypes defined using 5 loci in Figure 1A into
250 12 haplotypes (Figure 1D) which was consistent with another study that resolved Type X
251 *Toxoplasma* genotypes isolated from southern sea otters into 8 genotypes [26]. One isolate
252 each from subgroups A (n=4; 3168), C (n=4; 3160), D (n=3; 4166) and F (n=3; 3387) was
253 resolved further as a unique haplotype based on their inheritance of alleles at PK1, SAG4 and
254 GRA7 (Supplemental Table 1).
255
256 **Type X resembles a recombinant clade of F1 progeny from a natural cross**
257 eBURST is insufficiently resolved to distinguish among isolates that exist either as
258 variants within a clonal lineage, because they possess alleles that differ by only minor
259 mutational drift, or as genetic admixtures, that possess genetically distinct alleles derived
260 from different parental types. Previous studies suggested that Type X (or HG12) is a clonal
261 lineage infecting sea otters and wildlife in North America, and that HG12 represents an
262 expanded clone from a sexual cross between a Type II strain and a new genetic lineage, that
263 was referred to as "γ", but this was based largely on the alleles found at just two loci, GRA6
264 and GRA7 [24, 25, 27]. To investigate whether the genetic relationship among the 12 distinct
265 haplotypes is supported by a model of genetic drift, or that of recombination, individual
266 maximum likelihood trees were created for each of the 17 nuclear-encoded single gene copy
267 loci, the microsatellite locus, and the two organellar loci (Supplemental Figure 1). To
268 differentiate between alleles undergoing minor mutational drift, from those that have
269 evolved independently as distinct genetic outgroups, 1000 bootstrap replicates were run for
270 each tree, and supported nodes above 60% are indicated at each marker. For each tree, the
271 19 isolates identified as Type X were labeled in purple and isolates belonging to the clonal
272 lineage I (n=1), II (n=3), and III (n=1) were labeled in red, green, and blue respectively
273 (Figure 2).
274 At 5 genetic markers located on chromosomes Ib, II, III, XI, and XII, all 12 distinct Type
275 X haplotypes possessed a canonical Type II allele (green), with no minor mutational drift
276 detected, consistent with Type II being one of the parental genetic backgrounds
277 (Supplemental Table 1). At the L358 marker on chromosome V, however, all Type X isolates

278 except 3675 possessed a Type I or "α" allele (red), indicating the presence of mixed ancestry.
279 Further, at the BSR4 marker on chromosome IV, all Type X isolates except 3675 possessed
280 an entirely novel allele that was readily differentiated from Type I, II and III strains with
281 strong bootstrap support (Figure 2B). At all remaining loci, Type X isolates possessed either
282 a Type II or one of 2 genetically distinct alleles, which could not be readily explained by minor
283 mutational drift. Hence, across the 17 nuclear-encoded markers, each isolate possessed
284 either a Type II allele, or an allele of distinct ancestry that was referred to as belonging to the
285 "α" (red), "γ" (purple) or "δ" (orange) lineages.
286 The maternally inherited organellar genome markers were likewise mixed. At the
287 mitochondrial marker CO1, two alleles were identified, either an allele common to Type I, II
288 and III strains or a novel allele designated "γ" (Figure 2A). Further, 3 distinct alleles were
289 identified at the apicolast marker APICO: a Type I or "α" allele (red), a Type II allele (green)
290 or a novel "γ" allele (purple), supporting the mixed ancestry designation. Each of the 12 Type
291 X haplotypes were therefore a mosaic of mixed ancestry that possessed some combination
292 of Type II, Type I, "γ" or "δ" alleles that had segregated independently across the loci
293 investigated. The data are consistent with Type X existing as a recombinant clade resembling
294 F1 progeny that was recently derived, without sufficient time to develop minor mutational
295 drift (Supplemental Table 1).
296 In support of the sexual cross model, incongruity between phylogenies for each
297 haplotype was readily observed between markers located in different portions of the
298 genome. In Figure 2C, 8 isolates shared the same ancestry between the two genetic loci
299 located on separate chromosomes VIIa and VIIb (GRA7 and BAG1 respectively) and were
300 shaded either green (Type II alleles at both loci) or purple (γ lineage alleles at both loci). In
301 contrast, 13 isolates had independently segregated chromosomes of mixed ancestry, colored
302 orange and pink (to depict isolates that were discordant and possessed different ancestral
303 alleles at each locus), and the crossing of orange and pink lines highlighted this lack of
304 congruence (Figure 2C). For example, isolate 3671, had a γ lineage allele at GRA7 but a δ
305 lineage allele at BAG1. Furthermore, recombination within a chromosome was likewise
306 observed at linked markers for 18 strains (Figure 2D). For example, isolate 3387, had a γ
307 lineage allele at GRA7 but a δ lineage allele at SAG4, whereas 3675 had a II lineage allele
308 GRA7 but a γ lineage allele SAG4, formally establishing that recombination had occurred and
309 supporting genetic hybridization as the most plausible explanation for the origin of the 12
310 Type X haplotypes.
311
312 **Type X is an admixture cross between Type II and a novel genetic background by CGH-**
313 **array hybridization**
314 A subset of Type X isolates, 2 each from haplotypes A (3142, 3265), C (3026, 3045), F
315 (3503, 3429) and one each from D (3178) and H (3671) were next hybridized against a
316 photolithographic microarray that possessed 1517 polymorphic Type I, II and III strain-
317 specific genotyping probes distributed genome-wide [37]. This was done to test whether the
318 MLST genetic markers accurately predicted the presence of large genome-wide haplotype
319 specific blocks, consistent with genetic hybridization. Each Type-specific SNP on the CGH
320 array is represented as a dot colored either red, green or blue respectively for a hybridizing
321 strain that possesses a Type I, II or III SNP at the probe position. Dots colored grey only
322 identify probes that failed to hybridize, consistent with a strain that does not possess a Type

323 I, II or III specific SNP at the probe position. Hybridization with DNA from a Type I (GT1),
324 Type II (Me49), and Type III (CTG) strain identified hybridization patterns consistent with
325 genotype, indicating that all genotype-specific probes were functioning as expected (Figure
326 3A). Hybridization with the Type X isolates identified large, contiguous Type II haploblocks,
327 as evidenced by hybridization of all Type II-specific probes present on Chr III, VI, VIIb and
328 XII, for example. However, in other portions of the genome, the SNP diversity pattern was
329 novel. The hybridization pattern in these regions either shared patch-work similarity with
330 Type I (i.e., at Chr Ib, VI) or was highly divergent with large contiguous blocks of SNP probes
331 colored grey that failed to hybridize, which was consistent with introgression of a non-Type
332 I, II or III genetic ancestry, rather than minor mutational drift as the explanation for lack of
333 reactivity (i.e., Chr IV, XI, right end of Chr VII, XII, left end of Chr VIII).
334 The pattern of hybridizing SNPs was unique for each of the seven Type X isolates
335 examined, but it was only minimally different; either 3, 5, or 15 SNP differences were
336 detected in pairwise comparisons within haplotype A (3142, 3265), C (3026, 3045), and F
337 (3503, 3429), respectively (Figure 3A). At this resolution, it was not possible to distinguish
338 between CGH array hybridization efficiency, from that of minor mutational drift. Indeed, the
339 CGH arrays each had approximately the same number of SNP differences (3-15) between
340 haplotypes as they did within a haplotype, which is more consistent with hybridization
341 efficiency as the probable explanation for these hybridization differences. It did, however,
342 establish that 3671 is a distinct genetic admixture because it possessed a different Type II
343 hybridization pattern at Chr Ia, left hand of Chr Ib, VI, XI and right side of Chr XII, which was
344 distinct from all other Type X strains analyzed (Figure 3A).
345
346 **Type X is a recombinant clade at WGS resolution**
347 The lack of resolution using the CGH array approach indicated that WGS was required to
348 infer an accurate genetic history model for the Type X strains. Using genome-wide
349 polymorphism data derived by identifying all variant SNP positions after reference mapping
350 a Type I, II, III, 16 Type X, and three previously WGS sequenced HG12 isolates (WTD-1, RAY,
351 ARI) against the published Me49 genome, we generated an unrooted Neighbor-Net tree
352 (Figure 3B). The Neighbor-Net analysis established that all Type X isolates, in similar to Type
353 I and III strains, shared blocks of genome-wide ancestry with Type II, represented by the
354 reticulated pattern of edge blocks that depict recombination events between Type II and the
355 different genetic backgrounds within X, I and III. It did not support a clonal lineage
356 designation as the majority of Type X isolates appeared on separate branches (Figure 3B,
357 Inset). Only clonal complexes 1 and 4 (CC1, Type II; CC4, 3045, 3160, 4167) were supported
358 at WGS resolution, and 3 previously sequenced HG12 isolates from 2 people (ARI, RAY) and
359 a deer (WTD-1) formed well-supported clades with Type X isolates 4166 (WTD-1; RAY) and
360 3168 (ARI) indicating that Type X has a wider distribution beyond sea otter infection.
361 Outside of the reticulated network of edge blocks, the branch length for each Type X isolate
362 was significantly less than that observed for Type I and III, synonymous with a more recent
363 origin, without sufficient time to accumulate private SNPs by mutational drift. When murine
364 virulence data for each genotype was mapped onto the Neighbor-Net tree (see inset, Figure
365 3B), a clear partitioning of the different virulence phenotypes was resolved along branches
366 within the network, consistent with Type X isolates existing as sister progeny with only
367 minor mutational drift detected at each supported branch.

368 To identify the size, genome distribution, and total number of Type II admixed
369 haploblocks introgressed into each Type X isolate, pairwise SNP diversity plots were
370 generated for all Type X isolates, as well as reference Type I and III strains that have been
371 previously shown to have recombined with Type II. The Type I GT1 strain possessed an Me49
372 Chr Ia and IV, and an Me49 admixture block at the left side of Chr VIIa, and right end of Chr
373 XI (Figure 4). As expected, the Type III VEG strain possessed many more Me49 admixture
374 blocks distributed genome-wide than GT1, comprising ~40% of its genome. Only Type X
375 haplotype H (3671) possessed haploblocks that were highly similar in sequence to Me49, on
376 Chr Ia, II, III, VI, VIII, IX, XI, and XII (Figure 4). In other regions, however, 3671 shared regions
377 clearly Me49-like, but that appeared to have diverged somewhat by mutational drift
378 (average 3-5 SNPs per 10kb block) on Chr II or introgressed into Chr IV, V, VIIa, VIIb, VIII, IX,
379 and XII. Intermixed within the Type II regions of the 3671 genome were a limited number of
380 large haploblocks containing divergent SNP density synonymous with hybridization by a
381 strain possessing distinct γ or δ genetic ancestry (average of 50-100 SNPs per 10kb block).
382 When the pairwise SNP density analysis was expanded to different Type X haplotypes,
383 isolates within each haplotype possessed highly similar patchwork mosaics of Type II-like
384 or divergent (Type X) haploblocks that were specific to each haplotype. For example,
385 haplotype A isolates 3142 and 3168 SNP density plots were highly similar to each other, and
386 to haplotype C isolate 3045, but were readily distinguishable from isolates 4166 and 3503,
387 strains within haplotypes D and F, respectively (Figure 4). Specifically, isolate 3142 had a
388 Type X haploblock inheritance pattern at the right end of Chr V, and in the middle of Chr Ib,
389 VIIa, VIII, IX and XII whereas 4166, which was indistinguishable from a previously sequenced
390 HG12 strain (RAY) recovered from a human patient, possessed either Type II or 3671
391 haploblocks in these regions (Figure 4). Although the haplotype F isolate 3503 was highly
392 similar in genomic organization to 3142, it possessed a Type II haploblock at the right end of
393 Chr V that readily distinguished it from haplotype A. The pairwise SNP analysis established
394 that the majority of Type X haplotypes possess a genomic architecture that is strikingly
395 similar, but different at a limited number of admixture blocks. Coupled with the low allelic
396 diversity, the data support a genetic history model whereby Type X resembles a sexual clade
397 of recently-derived natural recombinants from a relatively limited number of crosses.
398
399 **PopNet analysis identifies only limited admixture blocks among Type X isolates**
400 While the Neighbor-Net analysis established that Type X exists as a recombinant clade of
401 strains, it failed to predict the precise number of genetic ancestries, or how they had admixed
402 positionally across the chromosomes. PopNet was used to paint chromosomes according to
403 their local inheritance patterns [38]. Included in the analysis were all sequenced Type X and
404 II strains, as well as reference Type I and III strains, known to have admixed with Type II.
405 PopNet identified 4 statistically supported ancestries and showed that each genome was a
406 mosaic of these distinct ancestries, supporting the admixture model.
407 Within the circle depicting the Type X clade, 5 distinct subgroupings were identified
408 based on the number and position of shared ancestral blocks; these were grouped together
409 based on line thickness (Figure 5A). These same groupings were supported both by Pairwise
410 SNP plots and Neighbor-Net analysis, but the PopNet analysis showed which of the 5
411 ancestries had introgressed positionally across the mosaic genomes. Clear recombination
412 blocks were readily resolved, and the genome architecture was remarkably similar between
413 the groupings. A custom script designed to identify only major crossover points between the

414 two parents (Type II and the mosaic ancestry of the γ/δ parent) identified either a limited
415 number of single and double recombination events, or the inheritance of whole
416 chromosomes of either Type II (Chr II) or γ/δ (Chr Ia, IV, XI) parental ancestry, consistent
417 with this group of strains representing related sister progeny (Figure 5B). Because murine
418 virulence plotted on the Neighbor-Net tree identified clear pathogenicity differences in mice
419 that clustered based on parasite genotype, and that the Type X strains resembled a natural
420 clade of recombinant progeny (akin to F_1) that possessed only a restricted number of
421 crossover points, a population-based QTL was performed to see if genes could be mapped
422 that contribute to murine virulence within the set of 18 natural isolates for which virulence
423 data existed (Figure 5C),
424
425 **Natural Population-based QTL identifies ROP33 (WNG-3) as a new murine Vir locus**
426 All Type X isolates recovered from sea otters resembled offspring from one (or a few
427 crosses) between a Type II strain, which is avirulent in mice, and unknown parent(s) that
428 are mosaic for two distinct ancestries. Because these natural isolates possessed a range of
429 virulence phenotypes when infected into mice (Figure 1C), a genome scan was performed to
430 determine the log-likelihood for association of discrete genome haploblocks with the acute
431 virulence phenotype. Four quantitative trait loci (QTL) peaks were identified with logarithm
432 of odds (LOD) scores 4.0 or greater on chromosomes V, VIIa, VIII, and X (Figure 6A). The
433 average size of the genomic regions spanned by the QTLs were in the range of 100-200kb,
434 except for the one on chromosome V, which was >700 kb (Figure 6B). To identify candidate
435 genes within the four peaks, the following inclusion criteria were assessed: presence of a
436 signal peptide and/or transmembrane domain, gene expression and polymorphism
437 differences, as well as every genes CRISPR genome-wide mutagenesis score for essentiality
438 [39](Supplemental Table 2). *ROP33* on chromosome VIIa stood out as the best candidate
439 gene to target for reverse genetics as it was abundantly expressed during acute infection, it
440 was predicted to be a functional serine-threonine protein kinase, it was highly polymorphic,
441 and is part of a family of divergent WNG (with-no-Gly-loop) kinases that regulate tubular
442 membrane biogenesis in the parasite [33]. Further, one allele of *ROP33* was strongly
443 correlated with acute virulence (p=0.00031; fishers two-sided exact test) (Figure 6C).
444 To determine whether *ROP33* was a novel murine virulence gene, the *ROP33* locus was
445 disrupted by targeted deletion using CRISPR-Cas9 facilitated double crossover homologous
446 recombination in the mouse virulent RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain [40]. This strain is
447 virulent in mice (LD_{100}=500 tachyzoites) and was engineered to accept targeted replacement
448 of the *rop33* gene using an HXGPRT gene flanked by 30bp of homology just outside of the
449 *rop33* promoter and 3'UTR region. Following selection in mycophenolic acid (MPA) and
450 xanthine (to select for the *HXGPRT* gene), the population was screened for disruption of the
451 *rop33* gene by PCR. Mice were next infected with either the parent RHΔ*ku80*Δ*rop18*Δ*hxgprt*
452 strain or the $Rop33^-$ mutant to assess murine virulence. Groups of 5 outbred CD-1 mice were
453 injected intraperitoneally with 500 tachyzoites and the results shown are for one of two
454 independent experiments. All mice infected with the parent RH line succumbed to infection
455 within 20 days (Figure 6D). In contrast, targeted deletion of the *rop33* gene was protective,
456 the majority of mice survived acute infection, and our results suggest that ROP33 is a novel
457 *Toxoplasma* acute virulence gene. Importantly, the lack of virulence was not the result of a
458 failure of the mutant parasite to proliferate *in vivo*, as no difference in parasite load was

459 detected through acute infection by bioluminescence imaging, indicating that deletion of
460 ROP33 does not appear to affect parasite proliferation *in vivo* (Figure 6E). All surviving mice
461 were confirmed to be positive by serology, indicating that they had been productively
462 infected. These data strongly suggest that ROP33 is a new virulence factor for murine
463 infection that was identified using natural isolates derived from a population genetic cross
464 performed in nature.
465
466 **Discussion**
467 Here, we investigated the emergence of a presumed clonal lineage of *Toxoplasma*
468 *gondii* (Type X; also referred to as HG12) recovered from a single marine host, infected
469 southern sea otters in California, that were collected longitudinally over a 7 year period in
470 order to determine the genetic basis for this epizootic in a federally listed, threatened
471 species. Although Type X has previously been identified as the 4th clonal lineage in North
472 American wildlife, work done here established that Type X is not a clonal lineage, but rather
473 a recombinant clade of strains that resemble F_1 progeny from at least one natural cross. DNA
474 sequence analysis using a wide selection of linked and unlinked markers across the nuclear
475 and organellar genomes of Type X isolates identified Type X to be a composite of Type II and
476 an unknown mosaic of two distinct ancestries, referred to as γ and δ, that have recombined
477 to produce a highly invasive clade of strains causing both subclinical infection as well as
478 mortality in threatened marine mammals, including the southern sea otter. Our data suggest
479 that Type X was derived from one or a limited number of crosses from two highly related
480 parental strains. One sea otter genotype was widely distributed and found to have expanded
481 to cause the majority of subclinical infections, but it was highly virulent to mice. Our natural
482 population dataset supports a model whereby *Toxoplasma*'s sexual life cycle is facilitating
483 the evolution and expansion of cryptically virulent strains that are capable of causing
484 pathogenic disease in a host-species specific manner. Hence, for generalist parasites that
485 have broad host ranges, natural selection among intermediate hosts appears to maximize
486 transmission and establishes how virulent strains can be maintained cryptically in nature.
487 The expansion of a mouse virulent clone into a new ecological niche, such as near-shore
488 marine mammals of the Eastern Pacific coastline, demonstrates how *Toxoplasma gondii* is
489 leveraging its intermediate host range to selectively partition parasite genetic diversity and
490 epidemic disease potential.
491 In previous studies, unknown strains having a distinct ancestry, referred to as α and
492 β, were found to have crossed with Type II to create the Type I and Type III clonal strains
493 respectively [11]. The frequency and pattern of recombined blocks inherited within the Type
494 X strains is parsimonious with previously described models for the genetic history of Types
495 I and III in which an unidentified ancestor (respectively α and β) sexually recombined with
496 an ancestral Type II to create the new lineages (Figure 5B) [11, 41]. Additionally, it has been
497 suggested previously that Type X is a product of genetic hybridization between a Type II
498 lineage and an unknown γ lineage [25, 27, 28, 32, 42]. Here, we have further refined this
499 perspective and concluded that Type X is a mosaic of a cross between Type II strains and two
500 distinct γ and δ ancestries . The γ/δ lineage has likely recombined at least once with Type II
501 to create the Type X recombinant clade of strains. Currently, no single isolate has been found
502 which could be defined as either γ or δ, in similar to the α and β lines thought to have admixed
503 with Type II to produce the clonal lines I and III, respectively [11, 41].

504 Previous work on Type X circulating in wildlife samples using a limited set of
505 genotyping markers identified Type X as the 4th clonal lineage in North America [10, 27].
506 However, this designation was based on the discovery of genetically distinct alleles at only
507 two loci, GRA6 and GRA7, which was the original basis for distinguishing the Type X lineage
508 as a new genotype from that of Type II, and that it had undergone a hybridization event with
509 a novel genotype [25]. It is now well established that Type X commonly infects wildlife across
510 North America [10, 14, 29, 30, 32, 36]. This study examined *Toxoplasma* strains collected
511 longitudinally across a 7 year period and large geographic range from sea otter infections,
512 and it was comprised of infections ranging from asymptomatic to subclinical or considered
513 the primary cause of death. As previous work had classified Type X as clonal, finding such a
514 myriad of different disease states among infected sea otters was not expected [10, 27, 32,
515 36]. To ascertain whether parasite genotype was a critical parameter influencing the
516 spectrum of disease, genotyping studies were carried out using an expanded set of markers
517 to ascertain the true population genetic structure of the isolates belonging to Type X. All
518 genotyping markers used in this study were sequenced to increase resolution rather than
519 evaluated strictly for their PCR-RFLP genotype, as was the standard for many of the previous
520 studies. While the markers used herein only surveyed a small part of the genome, they
521 identified at least 12 distinct haplotypes within Type X, confirming that they do not exist as
522 a clonal population, as hypothesized in another study [26]. Furthermore, at any given locus,
523 only a Type II allele or one or another of two distinct alleles was identified indicating limited
524 allelic diversity. Hence, each of the 12 haplotypes possessed one of three allelic types at any
525 one locus that had independently segregated across the markers examined. This result is
526 parsimonious with a recombinant clade of strains that reflect genetic hybrids from a limited,
527 but distinct set of ancestries. Our results were further confirmed both by CGH array analyses
528 as well as whole genome sequencing. The previous misclassification as a clonal population
529 appears to be the result of the low resolution genotyping analyses performed, and the
530 placement of the markers in predominantly Type II regions of the genome [10, 14, 43]. For
531 this reason, future studies should use WGS to not only discriminate between haplogroups
532 within a population but also to determine the true genetic ancestry of each haplogroup.
533 *Toxoplasma* population genetics would also greatly benefit from an increase in the number
534 of isolates that are whole genome sequenced. WGS should be expanded to include both a
535 greater diversity of strains as well as multiple isolates from within a canonical clonal lineage.
536 The Type I and III clonal lineages are thought to be derived from a limited number of
537 sexual crosses, and all strains within these clonotypes share large haploblocks of Type II-like
538 sequence [11, 42, 44, 45]. This finding also holds true for the Type X lineage. At all loci
539 surveyed, Type X displayed at least one isolate with a Type II inherited allele, and the
540 introgression of large haploblocks of Type II ancestry across the genomes was confirmed in
541 the 19 strains that were resolved at WGS resolution. In regions that did not clade with Type
542 II, two different allelic types were identified. We concluded that these represented distinct
543 ancestries that we referred to as γ and δ. This designation was supported by the inherent
544 reticulation of the clade, seen in the NeighborNet tree (Figure 2B), which possessed short
545 branch lengths radiating from a reticulated network with multiple strains on a single branch,
546 rather than producing a star-like phylogeny with multiple alleles present at each major
547 branch, the result of private SNPs that accumulated through genetic drift. In fact, the lack of
548 genetic drift between these loci supported genetic hybridization as the most parsimonious
549 explanation for the relationship between the Type X haplotypes.

550 Sexual replication with genetic recombination is known to occur in South America at high
551 frequency [41, 44, 46]. Additionally, recent studies on the population genetic structure of
552 *Toxoplasma* has established that genetic hybridization, at whole genome resolution, is extant
553 for the vast majority of sequenced strains that represent the breadth of genetic diversity
554 within *Toxoplasma* [15, 38]. Evidence of sexual recombination across the population, as well
555 as the evidence of sexual replication within the Type X clade shown here, may indicate that
556 sexual recombination is occurring more frequently than previously suspected within North
557 America.
558 No barriers appear to exist, or have been described, to limit sexual recombination from
559 occurring in the laboratory setting, although there is a lack of understanding for the genetic
560 factors that regulate sexual replication in the felid host [7, 45, 47-49]. It is not clear why
561 sexual recombination does not appear to be occurring in North American wildlife, to the
562 degree that it does in South America, although it is possible that self-mating within the
563 population masks the detection of genetic outcrossing [12, 42, 50]. Especially among closely
564 related isolates, unisexual mating cannot be resolved using current genotyping methods that
565 rely on low-resolution analyses and thus cannot readily distinguish asexual expansion from
566 unisexual mating [12].
567 *Toxoplasma* has been shown previously to utilize its sexual cycle to expand its biological
568 potential and alter its pathogenesis [7, 47, 49]. Sexual recombination, by the ability to
569 reshuffle parental alleles into new combinations, can produce progeny with a vast range of
570 altered biological potentials [7]. Our work herein has identified a single *Toxoplasma*
571 genotype that was derived from a recent genetic cross that has been naturally selected in
572 southern sea otters across a space of 7 years. Importantly, this same genotype was highly
573 pathogenic to outbred laboratory mice, with all mice dying in 10-15 days with an inoculum
574 of just 50 tachyzoites. In contrast, the two Type X B haplotype isolates that caused subclinical
575 infection in outbred laboratory mice were highly pathogenic to sea otters that stranded with
576 a fatal meningoencephalitis [25]. Our data thus support a virulence shift model for the
577 expansion and propagation of *Toxoplasma* in nature, whereby *Toxoplasma* utilizes its sexual
578 cycle to generate a wide spectrum of novel genotypes that possess altered biological
579 potentials that can be naturally selected across *Toxoplasma*'s vast host range for optimal
580 transmission. Hence, natural selection among intermediate hosts allows *Toxoplasma* to
581 maintain highly pathogenic strains cryptically that can cause serious disease in some animal
582 hosts, but not in others. Examples of this type of selection have been observed previously in
583 laboratory mice that express variable IRG (Immunity Related GTPase) gene arrays. IRGs are
584 host proteins primarily responsible for combatting *Toxoplasma* lysis of murine cells and
585 *Toxoplasma* encodes a suite of highly polymorphic Rhoptry kinase genes (ROPKs) whose
586 primary function is to inactivate host IRGs [51-53]. While laboratory mice are highly clonal,
587 wild mice and their subsequent IRGs are more diverse, as are the ROPKs expressed by
588 different *Toxoplasma* strains that are capable of infecting wild mice [17, 54]. Similarly, TLR
589 11/12 is one of the primary rodent innate immune response sensors (otherwise referred to
590 as PAMPs) that detects infection by *Toxoplasma*. However, not all intermediate hosts share
591 a functional combination of TLR11/12, thus, *Toxoplasma* strains that are selected for
592 infection in mice need to bypass TLR11/12 recognition, whereas this is not a barrier to
593 infection in a human cell, for example [54, 55]. Thus, intermediate hosts are capable of
594 naturally selecting *Toxoplasma* strain genotypes that establish chronic, transmissible
595 infections in a host species, parasite strain-dependent manner. In this way, the parasite finds

596 a balance between its virulence, fitness and infectious transmissibility across its wide host
597 range. It is this capacity to maintain cryptically virulent strains that may in part explain
598 *Toxoplasma's* success in nature; it has the ability to cause disease outbreaks, expand its host
599 range into new ecological niches, or alter its pathogenicity in both a parasite strain and host
600 genetic manner to maximize its transmission.

601 In other protozoan parasites, when host partitioning occurs, it is synonymous with
602 speciation. For example, *Plasmodium* parasites often partition by the mosquito host species
603 they co-evolve with to maximize parasite transmission [56-60]. Likewise, the apicomplexan
604 parasite *Sarcocystis* largely maintains separate species for each intermediate-definitive host
605 species combination to maintain its life cycle [12, 61]. In contrast, *Toxoplasma* forms
606 transmissible cysts in virtually all warm-blooded vertebrates, and these cysts are infectious
607 for its definitive felid host, as well as a large array of intermediate hosts. What is different, is
608 that across the genetic diversity of *Toxoplasma*, specific strain genotypes are being
609 selectively expanded among its vast array of different intermediate hosts, based primarily
610 on parasite genotype and the suite of polymorphic effector proteins encoded by a particular
611 strain. It is this fact that allows *Toxoplasma* to maintain cryptically virulent strains across its
612 broad intermediate host range. Hence, the intermediate host is playing a central role in the
613 natural selection, expansion, and maintenance of virulent strains across the broad host range
614 of this generalist parasite. In effect, infected hosts act as reservoirs for the pathogenic or
615 epidemic potential of the species. Our study showed that one sub-type of *Toxoplasma* had
616 expanded in the majority of sea otters because it caused primarily subclinical, chronic
617 infections in sea otters, but that it was uniformly lethal to outbred laboratory mice. This is
618 not unlike the expansion of strains bearing Type I alleles in birds that harbor asymptomatic
619 infections but prove highly pathogenic to laboratory mice [62]. Additionally, while Type II
620 isolates are commonly found infecting domestic livestock in North America, Type X is more
621 common in sylvatic hosts, leading others to propose that separate cycles exist within the
622 *Toxoplasma* population that overlap solely within the feline definitive host [10, 29, 63].

623 Finally, only a limited number of crossovers were detected among strains WGS
624 sequenced within the recombinant clade of Type X haplotypes. Because the pedigree of one
625 of the parents was known, we concluded that the isolates resembled F_1 progeny from a
626 natural cross, which prompted us to perform a population-based QTL analysis on an
627 unmanipulated, natural population of genetic hybrids. Taking advantage of a genome-wide,
628 high resolution SNP map we had generated for all sequenced haplotypes, our analysis
629 identified multiple punctate QTL peaks containing a limited number of candidate loci
630 associated with differences in mouse virulence for reverse genetic follow-up. The serine-
631 threonine protein kinase ROP33 stood out as the best candidate to influence pathogenicity,
632 it was identified previously to be an active protein kinase that is both abundantly expressed
633 and highly polymorphic. Further, ROP33 is a divergent WNG kinase (WNG-3) that is related
634 to ROP35 (WNG-1), a critical regulator of tubular membrane biogenesis, the formation of the
635 parasite intravesicular network (IVN) and the phosphorylation of dense granule proteins
636 associated with IVN biogenesis [33]. In fact dense granule proteins GRA2 and GRA6 are
637 required for IVN biogenesis, and IVN deficient parasites grow normally *in vitro*, but are
638 attenuated in mouse virulence assays [64, 65]. Other related ROP proteins, including ROP 5,
639 16, 17, and 18 are known to hijack host immune signaling pathways to alter *Toxoplasma*
640 virulence during rodent infection. However, all Type X isolates investigated in our study,
641 whether mouse virulent or avirulent, only expressed avirulent allele combinations of the

642 known ROP 5/16/18 virulence factors, which is why we chose to knock-out the gene in a
643 Ku80 deficient strain of *Toxoplasma* that was both mouse virulent, and deficient in ROP18.
644 This is because no change in mouse virulence was reported when ROP33 was deleted in a
645 wild-type RH strain background [66]. In addition, previous genetic mapping studies failed to
646 identify the QTL for this locus, even though it is polymorphic between Type I, II and III strains
647 for which genetic crosses have been performed. It is likely that the ROP5 and ROP18
648 virulence factors were dominant, and ROP33 is analogous to ROP17, another virulence factor
649 that was only identified after the virulence enhancing capacity of ROP18 was removed [40].
650 Our results show that parasites deficient in ROP33 in the highly permissive RHΔku80Δ*rop18*
651 parent, were avirulent, whereas an infectious dose of 500 tachyzoites of the RHΔku80Δ*rop18*
652 parent genotype was uniformly virulent. Specifically how ROP33 contributes to murine
653 virulence, what dense granule proteins it phosphorylates, the integrity of the IVN and what
654 host immune signaling pathways are modified are being actively investigated. Future work
655 will dissect the host or parasite factors that ROP33 targets to reduce or influence parasite
656 transmission and pathogenicity. Our study establishes that it is possible to use natural
657 isolates derived from a population genetic cross in nature to identify genetic loci associated
658 with a specific quantitative phenotype.
659
660 **Materials and Methods**
661
662 **Parasite culture**
663 Fifty-three previously published *T. gondii* strains isolated from brain and/or tissue
664 samples from stranded Pacific coast southern sea otters were provided by Dr. Patricia
665 Conrad [25, 26]. *Toxoplasma* strains were isolated at the Marine Wildlife Veterinary Care and
666 Research Centre in Santa Cruz, CA, or at the University of California, Davis School of
667 Veterinary Medicine. Cause of death (COD) was determined for each sea otter based on
668 primary pathology reports from marine veterinarians at date of necropsy, as previously
669 published [25, 26, 31, 67]. *Toxoplasma* isolates each had an ATOS ("as the otter swims")
670 number assigned to them based on the stranding location where each sea otter was
671 discovered. Isolates were named sequentially based on the date the sea otter host stranded.
672 Parasites were cultured and maintained in human foreskin fibroblast (HFF) monolayers as
673 described previously [68].
674
675 **DNA extraction and genetic typing markers**
676 Parasites were syringe lysed using a 27-gauge needle and filtered through a 3.0
677 micron polycarbonate filter to remove cellular debris. DNA was extracted from cell pellets
678 using the Qiagen DNeasy Blood and Tissue kit (Qiagen). All 53 isolates were typed using 5
679 previously described PCR-RFLP and one DNA sequencing marker [7, 25, 69-71]. BSR4, BAG1,
680 and ROP1 were PCR amplified and typed based on known RFLP identity. GRA6 and SAG3
681 were typed according to previously identified representative SNPs based on Sanger
682 sequence data done by the National Institute of Allergy and Infectious Disease's Rocky
683 Mountain Laboratory (NIAID RML).
684 Alleles observed via sequencing were designated as canonical I, II, or III. When
685 sequences differed from canonical alleles, previously described methods of characterization
686 were used [29]. Briefly, when genetic variation differed from references by two mutations
687 the allele was categorized as unique, while only one mutation difference was labeled a drifted

688 allele from the closest reference sequence. When more than two nucleotide mutations were
689 present, or the sequence was as distant from the reference strains as references are from
690 each other, alleles were classified as novel alleles.
691 21 strains representing the 8 unique genotypes identified by the original 5 markers
692 were selected to characterize further. These 21 strains were PCR amplified, sequenced, and
693 genotyped using 20 markers: 18 at both linked and unlinked genomic loci, encompassing 13
694 of the 14 chromosomes, and 2 loci on organellar genomes (apicoplast and mitochondria).
695 15,430 bp with 335 SNPs (~0.024% of the genome). Sequences were examined and
696 nucleotides verified using SeqMan Pro alignment software (Lasergene). Sequences for
697 reference strains (ME49, GT1, and VEG) were downloaded from ToxoDB [72].
698
699 **Generation of the *ROP33* knockout strain in *RHΔku80Δhxgprt* parasites**
700 Deletion of the *ROP33* gene was generated using the Type I parasite RH that was
701 rendered deficient in the expression of *KU80* and the drug selectable marker *HXGPRT*.
702 Briefly, the CRISPR/Cas9 targeting plasmid pSAG1:CAS9,U6:sgUPRT (received from Prof. L.
703 David Sibley, Department of Molecular Microbiology, Washington University School of
704 Medicine) was modified with a guide RNA (GTCGGACGCGAAACTCGCTT) to target the 5'
705 region of the *ROP33* gene (sgROP33) using Q5 mutagenesis (New England Biolabs, MA). Then
706 a CRISPR/Cas9 replacement construct was created using Gibson assembly (New England
707 Biolabs) to stitch together a 1kb flanking region of the 5' UTR region of ROP33
708 (TGGT1_chrVIIa, 3921773 to 3922773 bp) with the selectable marker HXGPRT and a 1kb
709 flanking region of the 3'UTR region of ROP33 (TGGT1_chrVIIa, 3926189 to 3927189 bp) to
710 replace the *ROP33* gene using the selectable drug cassette. To delete the *ROP33* gene, a total
711 of 50 µg of the sgROP33 and replacement construct (5:1 ratio) were co-transfected into a
712 ROP18 deficient strain of RH (RH *Δku80ΔhxgprtΔrop18*), which was cultured in human
713 foreskin fibroblasts (HFF) cells and maintained in Dulbecco's modified Eagle's medium
714 (DMEM) supplemented with 10% fetal bovine serum, 2 mM glutamine and 25 µg / ml
715 gentamicin. After transfection, parasites were selected for stable integration of the targeting
716 construct using mycophenolic acid (12.5 mg / mL in MeOH; MPA) / xanthine (20 mg / mL in
717 1M KOH; XAN) as described previously [73]. After 15 days, the resistant population was
718 cloned by limiting dilution and single clones were screened by PCR for targeted deletion of
719 the *ROP33* gene.
720
721 **Mouse virulence assay**
722 To determine the virulence of the sea otter isolates in a mouse model, groups of five
723 or more, 6-8 week-old, female CD1 outbred mice were intraperitoneally injected with 50 *T.*
724 *gondii* tachyzoites resuspended in 500 µl of PBS that had been expanded in HFF cells. Mice
725 were weighed daily to measure infection-induced cachexia and mouse survival was assayed
726 over 42 days [7]. At 14 days, mice were bled, and serum extracted to test for seroconversion
727 via indirect fluorescent antibody test (IFAT) against ME49 tachyzoites [74]. Strain virulence
728 was classified as follows: avirulent strains killed no seropositive mice within 42 days of
729 infection, intermediate virulent strains killed some but not all seropositive mice within 42
730 days post-infection, and virulent strains killed all seropositive mice. To assess the
731 contribution of ROP33 to mouse virulence, groups of five 6-8 week old female outbred CD-1
732 mice were injected intraperitoneally with 500 tachyzoites of either the parent

733 RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain or the *Rop33*$^{-}$ mutant. Infectivity and virulence were assessed
734 as above.
735
736 **Phylogenetic Tree and DNA Marker Analysis**
737 The default settings of Clustal X were used to align DNA sequences for all markers
738 individually and as a concatenated set of markers [75]. MSF files of aligned marker sequences
739 were imported into Molecular Evolutionary Genetic Analysis (MEGA) Version 7 to create a
740 maximum likelihood tree using Tamura-Nei model distance analysis with uniform rates of
741 substitution across all sites, and 1000 bootstrap support for all branch points [76].
742 Consensus trees were rooted on ME49, as Type II is an inherent parent in all strains analyzed.
743 All scales are by 0.001 nucleotide difference unit. Distinct parental lineages with bootstrap
744 support over 60% are indicated on each tree to distinguish II alleles from novel γ and δ
745 alleles. These novel alleles are indicated in purple and orange respectively on the 17 loci
746 MLST in Supplementary Table 2. The Nexus file of the concatenated marker alignment was
747 imported into SplitsTree 4 Version 4.13.1 where NetworkNet analysis was run using default
748 settings [77].
749
750 **eBURST Clonal Complex Analysis**
751 Alleles determined from the 17 nuclear encoded markers (excluding ROP1 due to its
752 classification as a microsatellite marker) were given numerical designations to create a
753 multilocus sequence-typing scheme for the eBURST program [78]. Default settings were
754 used to evaluate the isolated strains' alleles. Strains were displayed as dots and colors of dots
755 were depicted accordingly: Type I, II, III, and X being red, green, blue, and purple
756 respectively. Dot size was representative of the number of strains contained within each
757 genotype, larger dots representing more strains bearing the same genotype. Strains sharing
758 16 of the 17 nuclear encoded markers were designated as clonal complexes and represented
759 by connecting lines in the diagram shown.
760
761 **Genomic hybridization to Affymetrix arrays**
762 Genomic DNA was sheared and biotin labeled before hybridizing to a custom *T. gondii*
763 Affymetrix microarray as previously published [37]. High-fidelity SNPs were characterized
764 via a custom R script to identify SNPs belonging to each of the three reference strains (I, II,
765 and III). Three reference strains (GT1, ME49, and CTG) are shown to demonstrate ideal
766 hybridization within canonical lineages. Sea otter Type X DNA isolate hybridizations are
767 shown below the reference strains.
768
769 **Whole-Genome Sequence Analysis of Type X Recombination**
770 DNA from 16 isolates that grew best in tissue culture as well as 3 representative Type
771 X strains (ARI, RAY, and WTD1), for a total of 19 Type X strains was isolated. 3 μg of DNA
772 isolated from each of these 19 strains was sent to NIAID RML for whole-genome sequencing
773 using Illumina HiSeq technology. Fastq reads were reference mapped to the *T. gondii* ME49
774 assemblage Version 8.2 [72] using BWA 0.7.5a to align the reads to the reference genome
775 and GATK 3.7 in coordination with Picard 1.131 following best practices to quality control
776 the mapped reads [80]. Following mapping, the gVCF method of GATK was used to combine
777 SNP calls using stand_call_conf of 30.0, nct of 10, and ploidy of 1 (for haploid genomes) to

call 568,592 single nucleotide polymorphism positions across the whole genomes of these strains [80]. The derived VCF formatted SNP file was curated using GATK and VCFTools to produce a tabular file containing only biallelic SNPs with no large insertions or deletions [81]. A custom script was then utilized to convert this SNP file into a fasta file of strain polymorphic positions across the Type X and reference genomes. These SNP fasta files were run in SplitsTree4 using default parameters for BioNJ with 1000 bootstrap support to create a NeighborNet tree [72]. Interconnected reticulation between strains is indicative of recombination between strains while mitotic drift was visualized by divergent branching.

SNP Density Fingerprint Analysis of Recombinant Progeny

The same tabular, biallelic SNP file used in the NeighborNet analysis was modified using a custom R script based on location mapping to isolate strain-specific polymorphic locations from the tabular VCF. Further R scripts group these SNPs into 100 kb windows which are mapped across the genomes of these strains to display where diversity from the reference genome (ME49) is most apparent [44, 82]. The larger the number of SNPs in a particular window, the more divergent the strain is from the ME49 reference. Distinct haploblocks where genomic recombination has occurred are apparent in areas where haploblock diversity significantly varies from the surrounding regions on the same chromosome.

PopNet Characterization of Strain Interrelatedness

The tabular, biallelic SNP file created from WGS of Type X and used to derive the SNP density plots and NeighborNet tree was uploaded into PopNet using default parameters to assess the diversity and interrelatedness of strains [38]. Cytoscape was used to visualize the recombination and Markov clustering outputs. Genomes are displayed in circularized format with chromosomes concatenated into a circular genome display. The background of each strain is painted to match the group that shares the most common ancestry over the entire genome with the strain shown. Chromosome painting was done in 10 kb increments and the sequence haploblock was painted based on its shared ancestry. For instance, a Type X isolate haploblock that was most closely related to the ancestral Type II strain was painted green to indicate its inheritance. Strains that are more closely related have thicker connecting lines between the circles of the strain isolates.

Virulence Gene Identification by QTL

To find novel virulence alleles, the tabular, biallelic SNP file was utilized to run a quantitative trait locus (QTL) analysis on the Type X and Type II isolates of *Toxoplasma* sequenced at WGS resolution [83]. SNP calls were down-selected by a custom java script to include one SNP every 5 kb. This allows the QTL software to analyze the breadth of the WGS data in QTL, which was built to handle marker typing data without the depth inherent in WGS data. Custom scripts code the SNPs into reference (ME49) versus alternative (Type X's γ/δ lineage which substitute as the secondary parent) alleles. This down-selected marker dataset was combined with previously determined low-dose murine virulence for these strains. This data was then input into J/qtl and a one QTL genome scan was run using the default settings for the EM algorithm (maximum likelihood) with 1000 permutations to

822 identify genomic locations significantly associated with the percentage of murine death due
823 to Type X infection [84].
824 The QTL calculations identified four significantly associated regions based on a LOD
825 score of greater than 4.0. Within these genomic regions, 450 genes were predicted based
826 on ToxoDB documentation. These potential genes were down-selected based on presence
827 of a signal peptide and/or transmembrane domain, gene expression, polymorphism, and
828 genome-wide CRISPR score for essentiality to identify 32 virulence candidate genes within
829 the Type X strains. Of these, ROP33 was selected for further interrogation based on its high
830 LOD score and similarity to previously identified virulence effector proteins.
831
832 **Funding statement:**
833 This work was supported in part by the Intramural Research Program of the National
834 Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health and
835 by the National Science Foundation, Ecology of Infectious Diseases Grant 052576.
836
837 **Ethics statements:**
838 The animal study protocol LPD 22E was reviewed and approved by the Animal Care and Use
839 Committee of the Intramural Research Program of the National Institute of Allergy and
840 Infectious Diseases, National Institutes of Health. No human studies or identifiable human
841 data are presented in this study. All datasets generated for this study are available upon
842 request to the corresponding author.
843
844 **Acknowledgements**
845 We thank Andrea Packham and Ann Melli for their assistance in culturing and isolation of
846 the *Toxoplasma* isolates, and Javi Zhang for his help with PopNet analysis. We also thank all
847 volunteers and staff at the California Department of Fish and Wildlife, the Monterey Bay
848 Aquarium, the Marine Mammal Center, and the United States Geological Survey for their
849 efforts to recover stranded sea otters that facilitated the isolation of *Toxoplasma gondii*
850 strains utilized in this paper.
851
852 **References**
853 1. Vijaykrishna D, Smith GJ, Pybus OG, Zhu H, Bhatt S, Poon LL, et al. Long-term evolution
854 and transmission dynamics of swine influenza A virus. Nature. 2011;473(7348):519-22.
855 doi: 10.1038/nature10004. PubMed PMID: 21614079.
856 2. Bellanger X, Payot S, Leblond-Bourget N, Guedon G. Conjugative and mobilizable
857 genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev.
858 2014;38(4):720-60. doi: 10.1111/1574-6976.12058. PubMed PMID: 24372381.
859 3. Schneider G, Dobrindt U, Middendorf B, Hochhut B, Szijarto V, Emody L, et al.
860 Mobilisation and remobilisation of a large archetypal pathogenicity island of
861 uropathogenic *Escherichia coli* in vitro support the role of conjugation for horizontal
862 transfer of genomic islands. BMC Microbiol. 2011;11:210. doi: 10.1186/1471-2180-11-
863 210. PubMed PMID: 21943043; PubMed Central PMCID: PMC3202238.
864 4. Wardal E, Markowska K, Zabicka D, Wroblewska M, Giemza M, Mik E, et al. Molecular
865 analysis of vanA outbreak of *Enterococcus faecium* in two Warsaw hospitals: the
866 importance of mobile genetic elements. Biomed Res Int. 2014;2014:575367. doi:

867 10.1155/2014/575367. PubMed PMID: 25003118; PubMed Central PMCID:
868 PMC4070583.
869 5. English ED, Adomako-Ankomah Y, Boyle JP. Secreted effectors in *Toxoplasma gondii* and
870 related species: determinants of host range and pathogenesis? Parasite immunology.
871 2015;37(3):127-40. doi: 10.1111/pim.12166. PubMed PMID: 25655311.
872 6. Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, et al. Same-sex
873 mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature.
874 2005;437(7063):1360-4. Epub 2005/10/14. doi: 10.1038/nature04220. PubMed
875 PMID: 16222245.
876 7. Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC. Success and virulence in
877 *Toxoplasma* as the result of sexual recombination between two distinct ancestries.
878 Science. 2001;294(5540):161-5. doi: DOI 10.1126/science.1061888. PubMed PMID:
879 WOS:000171448800052.
880 8. Boothroyd JC, Grigg ME. Population biology of *Toxoplasma gondii* and its relevance to
881 human infection: do different strains cause different disease? Curr Opin Microbiol. 5.
882 England2002. p. 438-42.
883 9. Boyer K, Hill D, Mui E, Wroblewski K, Karrison T, Dubey JP, et al. Unrecognized ingestion
884 of *Toxoplasma gondii* oocysts leads to congenital toxoplasmosis and causes epidemics in
885 North America. Clin Infect Dis. 2011;53(11):1081-9. Epub 2011/10/25. doi:
886 10.1093/cid/cir667. PubMed PMID: 22021924; PubMed Central PMCID:
887 PMCPMC3246875.
888 10. Behnke MS, Khan A, Wootton JC, Dubey JP, Tang K, Sibley LD. Virulence differences in
889 *Toxoplasma* mediated by amplification of a family of polymorphic pseudokinases. Proc
890 Natl Acad Sci U S A. 2011;108(23):9631-6. doi: 10.1073/pnas.1015338108. PubMed
891 PMID: 21586633; PubMed Central PMCID: PMC3111276.
892 11. Boyle JP, Rajasekar B, Saeij JP, Ajioka JW, Berriman M, Paulsen I, et al. Just one cross
893 appears capable of dramatically altering the population biology of a eukaryotic
894 pathogen like *Toxoplasma gondii*. Proc Natl Acad Sci U S A. 2006;103(27):10514-9. doi:
895 10.1073/pnas.0510319103. PubMed PMID: 16801557; PubMed Central PMCID:
896 PMC1502489.
897 12. Wendte JM, Miller MA, Lambourn DM, Magargal SL, Jessup DA, Grigg ME. Self-mating in
898 the definitive host potentiates clonal outbreaks of the apicomplexan parasites
899 *Sarcocystis neurona* and *Toxoplasma gondii*. PLoS Genetics. 2010;6(12):e1001261. doi:
900 10.1371/journal.pgen.1001261. PubMed PMID: 21203443; PubMed Central PMCID:
901 PMC3009688.
902 13. Boothroyd JC. Expansion of host range as a driving force in the evolution of *Toxoplasma*.
903 Mem I Oswaldo Cruz. 2009;104(2):179-84. PubMed PMID: WOS:000267051500009.
904 14. Howe DK, Sibley LD. *Toxoplasma gondii* comprises three clonal lineages: correlation of
905 parasite genotype with human disease. J Infect Dis. 1995;172(6):1561-6. PubMed PMID:
906 7594717.
907 15. Lorenzi H, Khan A, Behnke MS, Namasivayam S, Swapna LS, Hadjithomas M, et al. Local
908 admixture of amplified and diversified secreted pathogenesis determinants shapes
909 mosaic *Toxoplasma gondii* genomes. Nat Commun. 2016;7:10147. doi:
910 10.1038/ncomms10147. PubMed PMID: 26738725; PubMed Central PMCID:
911 PMCPMC4729833.

16. Sibley LD, Boothroyd JC. Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. Nature. 1992;359(6390):82-5. doi: 10.1038/359082a0. PubMed PMID: 1355855.
17. Lilue J, Muller UB, Steinfeldt T, Howard JC. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. Elife. 2013;2:e01298. doi: 10.7554/eLife.01298. PubMed PMID: 24175088; PubMed Central PMCID: PMC3810784.
18. Steinfeldt T, Konen-Waisman S, Tong L, Pawlowski N, Lamkemeyer T, Sibley LD, et al. Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent *Toxoplasma gondii*. PLoS Biology. 2010;8(12):e1000576. Epub 2011/01/05. doi: 10.1371/journal.pbio.1000576. PubMed PMID: 21203588; PubMed Central PMCID: PMCPMC3006384.
19. Cavailles P, Sergent V, Bisanz C, Papapietro O, Colacios C, Mas M, et al. The rat *Toxo1* locus directs toxoplasmosis outcome and controls parasite proliferation and spreading by macrophage-dependent mechanisms. Proc Natl Acad Sci U S A. 2006;103(3):744-9. Epub 2006/01/13. doi: 10.1073/pnas.0506643103. PubMed PMID: 16407112; PubMed Central PMCID: PMCPMC1334643.
20. Cirelli KM, Gorfu G, Hassan MA, Printz M, Crown D, Leppla SH, et al. Inflammasome sensor NLRP1 controls rat macrophage susceptibility to *Toxoplasma gondii.* PLoS Pathog. 2014;10(3):e1003927. Epub 2014/03/15. doi: 10.1371/journal.ppat.1003927. PubMed PMID: 24626226; PubMed Central PMCID: PMCPMC3953412.
21. Ewald SE, Chavarria-Smith J, Boothroyd JC. NLRP1 is an inflammasome sensor for *Toxoplasma gondii*. Infection and Immunity. 2014;82(1):460-8. Epub 2013/11/13. doi: 10.1128/IAI.01170-13. PubMed PMID: 24218483; PubMed Central PMCID: PMCPMC3911858.
22. Gorfu G, Cirelli KM, Melo MB, Mayer-Barber K, Crown D, Koller BH, et al. Dual role for inflammasome sensors NLRP1 and NLRP3 in murine resistance to *Toxoplasma gondii*. mBio. 2014;5(1). Epub 2014/02/20. doi: 10.1128/mBio.01117-13. PubMed PMID: 24549849; PubMed Central PMCID: PMCPMC3944820.
23. Dubey JP, Shen SK, Kwok OC, Frenkel JK. Infection and immunity with the RH strain of *Toxoplasma gondii* in rats and mice. The Journal of Parasitology. 1999;85(4):657-62. PubMed PMID: 10461945.
24. Conrad PA, Miller MA, Kreuder C, James ER, Mazet J, Dabritz H, et al. Transmission of *Toxoplasma*: clues from the study of sea otters as sentinels of *Toxoplasma gondii* flow into the marine environment. Int J Parasitol. 2005;35(11-12):1155-68. Epub 2005/09/15. doi: 10.1016/j.ijpara.2005.07.002. PubMed PMID: 16157341.
25. Miller MA, Grigg ME, Kreuder C, James ER, Melli AC, Crosbie PR, et al. An unusual genotype of *Toxoplasma gondii* is common in California sea otters (*Enhydra lutris nereis*) and is a cause of mortality. Int J Parasitol. 2004;34(3):275-84. doi: 10.1016/j.ijpara.2003.12.008. PubMed PMID: 15003489.
26. Shapiro K, VanWormer E, Packham A, Dodd E, Conrad PA, Miller M. Type X strains of *Toxoplasma gondii* are virulent for southern sea otters (*Enhydra lutris nereis*) and present in felids from nearby watersheds. Proc Biol Sci. 2019;286(1909):20191334. Epub 2019/08/23. doi: 10.1098/rspb.2019.1334. PubMed PMID: 31431162; PubMed Central PMCID: PMCPMC6732395.

27. Khan A, Dubey JP, Su C, Ajioka JW, Rosenthal BM, Sibley LD. Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. Int J Parasitol. 2011;41(6):645-55. doi: 10.1016/j.ijpara.2011.01.005. PubMed PMID: 21320505; PubMed Central PMCID: PMC3081397.
28. Dubey JP, Velmurugan GV, Rajendran C, Yabsley MJ, Thomas NJ, Beckmen KB, et al. Genetic characterisation of *Toxoplasma gondii* in wildlife from North America revealed widespread and high prevalence of the fourth clonal type. Int J Parasitol. 2011;41(11):1139-47. doi: 10.1016/j.ijpara.2011.06.005. PubMed PMID: 21802422.
29. VanWormer E, Miller MA, Conrad PA, Grigg ME, Rejmanek D, Carpenter TE, et al. Using molecular epidemiology to track *Toxoplasma gondii* from terrestrial carnivores to marine hosts: implications for public health and conservation. PLoS Neglected Tropical Diseases. 2014;8(5):e2852. doi: 10.1371/journal.pntd.0002852. PubMed PMID: 24874796; PubMed Central PMCID: PMC4038486.
30. Gibson AK, Raverty S, Lambourn DM, Huggins J, Magargal SL, Grigg ME. Polyparasitism is associated with increased disease severity in *Toxoplasma gondii*-infected marine sentinel species. PLoS Neglected Tropical Diseases. 2011;5(5):e1142. doi: 10.1371/journal.pntd.0001142. PubMed PMID: 21629726; PubMed Central PMCID: PMC3101184.
31. Kreuder C, Miller MA, Jessup DA, Lowenstine LJ, Harris MD, Ames JA, et al. Patterns of mortality in southern sea otters (*Enhydra lutris nereis*) from 1998-2001. J Wildl Dis. 2003;39(3):495-509. doi: 10.7589/0090-3558-39.3.495. PubMed PMID: 14567210.
32. Sundar N, Cole RA, Thomas NJ, Majumdar D, Dubey JP, Su C. Genetic diversity among sea otter isolates of *Toxoplasma gondii.* Veterinary Parasitology. 2008;151(2-4):125-32. doi: 10.1016/j.vetpar.2007.11.012. PubMed PMID: 18155841.
33. Beraki T, Hu X, Broncel M, Young JC, O'Shaughnessy WJ, Borek D, et al. Divergent kinase regulates membrane ultrastructure of the *Toxoplasma* parasitophorous vacuole. Proc Natl Acad Sci U S A. 2019;116(13):6361-70. Epub 2019/03/10. doi: 10.1073/pnas.1816161116. PubMed PMID: 30850550; PubMed Central PMCID: PMCPMC6442604.
34. Miller MA, Miller WA, Conrad PA, James ER, Melli AC, Leutenegger CM, et al. Type X *Toxoplasma gondii* in a wild mussel and terrestrial carnivores from coastal California: new linkages between terrestrial mammals, runoff and toxoplasmosis of sea otters. Int J Parasitol. 2008;38(11):1319-28. doi: 10.1016/j.ijpara.2008.02.005. PubMed PMID: 18452923.
35. Saeij JP, Boyle JP, Boothroyd JC. Differences among the three major strains of *Toxoplasma gondii* and their specific interactions with the infected host. Trends in Parasitology. 2005;21(10):476-81. doi: 10.1016/j.pt.2005.08.001. PubMed PMID: 16098810.
36. Su C, Khan A, Zhou P, Majumdar D, Ajzenberg D, Darde ML, et al. Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. Proc Natl Acad Sci U S A. 2012;109(15):5844-9. doi: 10.1073/pnas.1203190109. PubMed PMID: 22431627; PubMed Central PMCID: PMCPMC3326454.
37. Khan A, Miller N, Roos DS, Dubey JP, Ajzenberg D, Darde ML, et al. A monomorphic haplotype of chromosome Ia is associated with widespread success in clonal and nonclonal populations of *Toxoplasma gondii*. mBio. 2011;2(6):e00228-11. doi:

1003 10.1128/mBio.00228-11. PubMed PMID: 22068979; PubMed Central PMCID:
1004 PMC3215432.
1005 38. Zhang J, Khan A, Kennard A, Grigg ME, Parkinson J. PopNet: A Markov Clustering
1006 Approach to Study Population Genetic Structure. Mol Biol Evol. 2017;34(7):1799-811.
1007 Epub 2017/04/07. doi: 10.1093/molbev/msx110. PubMed PMID: 28383661; PubMed
1008 Central PMCID: PMCPMC5850731.
1009 39. Sidik SM, Huet D, Ganesan SM, Huynh MH, Wang T, Nasamu AS, et al. A Genome-wide
1010 CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes. Cell.
1011 2016;166(6):1423-35 e12. Epub 2016/09/07. doi: 10.1016/j.cell.2016.08.019. PubMed
1012 PMID: 27594426; PubMed Central PMCID: PMCPMC5017925.
1013 40. Etheridge RD, Alaganan A, Tang K, Lou HJ, Turk BE, Sibley LD. The *Toxoplasma*
1014 pseudokinase ROP5 forms complexes with ROP18 and ROP17 kinases that synergize to
1015 control acute virulence in mice. Cell Host & Microbe. 2014;15(5):537-50. Epub
1016 2014/05/17. doi: 10.1016/j.chom.2014.04.002. PubMed PMID: 24832449; PubMed
1017 Central PMCID: PMCPMC4086214.
1018 41. Khan A, Fux B, Su C, Dubey JP, Darde ML, Ajioka JW, et al. Recent transcontinental sweep
1019 of *Toxoplasma gondii* driven by a single monomorphic chromosome. Proc Natl Acad Sci
1020 U S A. 2007;104(37):14872-7. Epub 2007/09/07. doi: 10.1073/pnas.0702356104.
1021 PubMed PMID: 17804804; PubMed Central PMCID: PMCPMC1965483.
1022 42. Grigg ME, Sundar N. Sexual recombination punctuated by outbreaks and clonal
1023 expansions predicts *Toxoplasma gondii* population genetics. Int J Parasitol.
1024 2009;39(8):925-33. doi: 10.1016/j.ijpara.2009.02.005. PubMed PMID: 19217909;
1025 PubMed Central PMCID: PMC2713429.
1026 43. Ajzenberg D, Cogne N, Paris L, Bessieres MH, Thulliez P, Filisetti D, et al. Genotype of 86
1027 *Toxoplasma gondii* isolates associated with human congenital toxoplasmosis, and
1028 correlation with clinical findings. J Infect Dis. 2002;186(5):684-9. Epub 2002/08/27.
1029 doi: 10.1086/342663. PubMed PMID: 12195356.
1030 44. Minot S, Melo MB, Li F, Lu D, Niedelman W, Levine SS, et al. Admixture and
1031 recombination among *Toxoplasma gondii* lineages explain global genome diversity. P
1032 Natl Acad Sci USA. 2012;109(33):13458-63. doi: DOI 10.1073/pnas.1117047109.
1033 PubMed PMID: WOS:000307807000067; PubMed Central PMCID: PMCPMC3421188.
1034 45. Khan A, Shaik JS, Behnke M, Wang QL, Dubey JP, Lorenzi HA, et al. NextGen sequencing
1035 reveals short double crossovers contribute disproportionately to genetic diversity in
1036 *Toxoplasma gondii*. Bmc Genomics. 2014;15. doi: Artn 1168
1037 10.1186/1471-2164-15-1168. PubMed PMID: WOS:000349049000003.
1038 46. Rajendran C, Su C, Dubey JP. Molecular genotyping of *Toxoplasma gondii* from Central
1039 and South America revealed high diversity within and between populations. Infection,
1040 Genetics and Evolution : journal of molecular epidemiology and evolutionary genetics in
1041 infectious diseases. 2012;12(2):359-68. Epub 2012/01/10. doi:
1042 10.1016/j.meegid.2011.12.010. PubMed PMID: 22226702.
1043 47. Saeij JP, Boyle JP, Coller S, Taylor S, Sibley LD, Brooke-Powell ET, et al. Polymorphic
1044 secreted kinases are key virulence factors in toxoplasmosis. Science.
1045 2006;314(5806):1780-3. doi: 10.1126/science.1133690. PubMed PMID: 17170306;
1046 PubMed Central PMCID: PMC2646183.
1047 48. Saeij JP, Coller S, Boyle JP, Jerome ME, White MW, Boothroyd JC. *Toxoplasma* co-opts host
1048 gene expression by injection of a polymorphic kinase homologue. Nature.

1049 2007;445(7125):324-7. doi: 10.1038/nature05395. PubMed PMID: 17183270; PubMed
1050 Central PMCID: PMC2637441.
1051 49. Taylor S, Barragan A, Su C, Fux B, Fentress SJ, Tang K, et al. A secreted serine-threonine
1052 kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii.* Science.
1053 2006;314(5806):1776-80. doi: 10.1126/science.1133643. PubMed PMID: 17170305.
1054 50. Ferguson D. *Toxoplasma gondii* and sex: essential or optional extra? Trends in
1055 Parasitology. 2002;18(8):351. Epub 2002/10/16. PubMed PMID: 12377284.
1056 51. Fleckenstein MC, Reese ML, Konen-Waisman S, Boothroyd JC, Howard JC, Steinfeldt T. A
1057 *Toxoplasma gondii* pseudokinase inhibits host IRG resistance proteins. PLoS Biology.
1058 2012;10(7):e1001358. doi: 10.1371/journal.pbio.1001358. PubMed PMID: 22802726;
1059 PubMed Central PMCID: PMC3393671.
1060 52. Hunn JP, Feng CG, Sher A, Howard JC. The immunity-related GTPases in mammals: a fast-
1061 evolving cell-autonomous resistance system against intracellular pathogens. Mamm
1062 Genome. 2011;22(1-2):43-54. doi: 10.1007/s00335-010-9293-3. PubMed PMID:
1063 21052678; PubMed Central PMCID: PMC3438224.
1064 53. Yamamoto M, Okuyama M, Ma JS, Kimura T, Kamiyama N, Saiga H, et al. A cluster of
1065 interferon-gamma-inducible p65 GTPases plays a critical role in host defense against
1066 *Toxoplasma gondii.* Immunity. 2012;37(2):302-13. doi: 10.1016/j.immuni.2012.06.009.
1067 PubMed PMID: 22795875.
1068 54. Gazzinelli RT, Mendonca-Neto R, Lilue J, Howard J, Sher A. Innate Resistance against
1069 *Toxoplasma gondii*: An Evolutionary Tale of Mice, Cats, and Men. Cell Host & Microbe.
1070 2014;15(2):132-8. Epub 2014/02/18. doi: 10.1016/j.chom.2014.01.004. PubMed
1071 PMID: 24528860.
1072 55. Koblansky AA, Jankovic D, Oh H, Hieny S, Sungnak W, Mathur R, et al. Recognition of
1073 profilin by Toll-like receptor 12 is critical for host resistance to *Toxoplasma gondii.*
1074 Immunity. 2013;38(1):119-30. doi: 10.1016/j.immuni.2012.09.016. PubMed PMID:
1075 23246311; PubMed Central PMCID: PMC3601573.
1076 56. Mackinnon MJ, Read AF. Virulence in malaria: an evolutionary viewpoint. Philosophical
1077 transactions of the Royal Society of London Series B, Biological sciences.
1078 2004;359(1446):965-86. doi: 10.1098/rstb.2003.1414. PubMed PMID: 15306410;
1079 PubMed Central PMCID: PMC1693375.
1080 57. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, et al. A molecular
1081 marker of artemisinin-resistant *Plasmodium falciparum* malaria. Nature.
1082 2014;505(7481):50-5. Epub 2013/12/20. doi: 10.1038/nature12876. PubMed PMID:
1083 24352242.
1084 58. Bopp SE, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, et al. Mitotic evolution
1085 of *Plasmodium falciparum* shows a stable core genome but recombination in antigen
1086 families. PLoS Genetics. 2013;9(2):e1003293. doi: 10.1371/journal.pgen.1003293.
1087 PubMed PMID: 23408914; PubMed Central PMCID: PMC3567157.
1088 59. Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, et al. Multiple
1089 populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. Nature
1090 Genetics. 2013;45(6):648-55. Epub 2013/04/30. doi: 10.1038/ng.2624. PubMed PMID:
1091 23624527; PubMed Central PMCID: PMC3807790.
1092 60. Arisue N, Hashimoto T. Phylogeny and evolution of apicoplasts and apicomplexan
1093 parasites. Parasitology International. 2014. doi: 10.1016/j.parint.2014.10.005. PubMed
1094 PMID: 25451217.

1095 61. Barbosa L, Johnson CK, Lambourn DM, Gibson AK, Haman KH, Huggins JL, et al. A novel
1096 *Sarcocystis neurona* genotype XIII is associated with severe encephalitis in an
1097 unexpectedly broad range of marine mammals from the northeastern Pacific Ocean. Int
1098 J Parasitol. 2015. doi: 10.1016/j.ijpara.2015.02.013. PubMed PMID: 25997588.
1099 62. Dubey JP, Graham DH, da Silva DS, Lehmann T, Bahia-Oliveira LM. *Toxoplasma gondii*
1100 isolates of free-ranging chickens from Rio de Janeiro, Brazil: mouse mortality, genotype,
1101 and oocyst shedding by cats. The Journal of Parasitology. 2003;89(4):851-3. Epub
1102 2003/10/10. doi: 10.1645/GE-60R. PubMed PMID: 14533703.
1103 63. Wendte JM, Gibson AK, Grigg ME. Population genetics of *Toxoplasma gondii*: new
1104 perspectives from parasite genotypes in wildlife. Veterinary Parasitology.
1105 2011;182(1):96-111. doi: 10.1016/j.vetpar.2011.07.018. PubMed PMID: 21824730.
1106 64. Labruyere E, Lingnau M, Mercier C, Sibley LD. Differential membrane targeting of the
1107 secretory proteins GRA4 and GRA6 within the parasitophorous vacuole formed by
1108 *Toxoplasma gondii*. Mol Biochem Parasitol. 1999;102(2):311-24. Epub 1999/09/25. doi:
1109 10.1016/s0166-6851(99)00092-4. PubMed PMID: 10498186.
1110 65. Mercier C, Howe DK, Mordue D, Lingnau M, Sibley LD. Targeted disruption of the GRA2
1111 locus in *Toxoplasma gondii* decreases acute virulence in mice. Infection and Immunity.
1112 1998;66(9):4176-82. Epub 1998/08/26. PubMed PMID: 9712765.
1113 66. Wang JL, Li TT, Elsheikha HM, Chen K, Zhu WN, Yue DM, et al. Functional
1114 Characterization of Rhoptry Kinome in the Virulent *Toxoplasma gondii* RH Strain. Front
1115 Microbiol. 2017;8:84. Epub 2017/02/09. doi: 10.3389/fmicb.2017.00084. PubMed
1116 PMID: 28174572; PubMed Central PMCID: PMCPMC5258691.
1117 67. Miller MA, Gardner IA, Packham A, Mazet JK, Hanni KD, Jessup D, et al. Evaluation of an
1118 Indirect Fluorescent Antibody Test (Ifat) for Demonstration of Antibodies to
1119 *Toxoplasma Gondii* in the Sea Otter (*Enhydra Lutris*). Journal of Parasitology.
1120 2002;88(3):594-9. doi: 10.1645/0022-3395(2002)088[0594:eoaifa]2.0.co;2.
1121 68. Pszenny V, Angel SO, Duschak VG, Paulino M, Ledesma B, Yabo MI, et al. Molecular
1122 cloning, sequencing and expression of a serine proteinase inhibitor gene from
1123 *Toxoplasma gondii*. Mol Biochem Parasitol. 2000;107(2):241-9. PubMed PMID:
1124 10779600.
1125 69. Su C, Zhang X, Dubey JP. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP
1126 markers: a high resolution and simple method for identification of parasites. Int J
1127 Parasitol. 2006;36(7):841-8. doi: 10.1016/j.ijpara.2006.03.003. PubMed PMID:
1128 16643922.
1129 70. Fazaeli A, Carter PE, Darde ML, Pennington TH. Molecular typing of *Toxoplasma gondii*
1130 strains by GRA6 gene sequence analysis. Int J Parasitol. 2000;30(5):637-42. PubMed
1131 PMID: 10779578.
1132 71. Howe DK, Sibley LD. *Toxoplasma gondii*: analysis of different laboratory stocks of the RH
1133 strain reveals genetic heterogeneity. Exp Parasitol. 1994;78(2):242-5. doi:
1134 10.1006/expr.1994.1024. PubMed PMID: 7907030.
1135 72. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, et al. ToxoDB: an integrated
1136 *Toxoplasma gondii* database resource. Nucleic Acids Res. 2008;36(Database
1137 issue):D553-6. doi: 10.1093/nar/gkm981. PubMed PMID: 18003657.
1138 73. Donald RG, Carter D, Ullman B, Roos DS. Insertional tagging, cloning, and expression of
1139 the *Toxoplasma gondii* hypoxanthine-xanthine-guanine phosphoribosyltransferase
1140 gene. Use as a selectable marker for stable transformation. J Biol Chem.

1996;271(24):14010-9. Epub 1996/06/14. doi: 10.1074/jbc.271.24.14010. PubMed PMID: 8662859.
74. Fletcher S. Indirect Fluorescent Antibody Technique in the Serology of *Toxoplasma gondii*. J Clin Pathol. 1965;18:193-9. Epub 1965/03/01. PubMed PMID: 14276154.
75. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8. doi: 10.1093/bioinformatics/btm404. PubMed PMID: 17846036.
76. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. doi: 10.1093/molbev/mst197. PubMed PMID: 24132122.
77. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67. doi: 10.1093/molbev/msj030. PubMed PMID: 16221896.
78. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. Journal of Bacteriology. 2004;186(5):1518-30. doi: 10.1128/jb.186.5.1518-1530.2004.
79. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19(18):2496-7. doi: 10.1093/bioinformatics/btg359.
80. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;11(1110):11 0 1- 0 33. doi: 10.1002/0471250953.bi1110s43. PubMed PMID: 25431634.
81. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330.
82. Yin, T., Cook, D., & Lawrence, M. (2012). ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol, 13*(8), R77. doi:10.1186/gb-2012-13-8-r77
83. Behnke, M. S., Khan, A., Lauron, E. J., Jimah, J. R., Wang, Q., Tolia, N. H., & Sibley, L. D. (2015). Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of Toxoplasma gondii. *PLoS Genet, 11*(8), e1005434. doi:10.1371/journal.pgen.1005434.
84. Smith, R., Sheppard, K., DiPetrillo, K., & Churchill, G. (2009). Quantitative trait locus analysis using J/qtl. *Methods Mol Biol, 573*, 175-188. doi:10.1007/978-1-60761-247-6_10.

Competing interests

The authors declare no competing financial interests.

Figure Legends

<u>Figure 1</u>: Parasite diversity and mouse virulence within the Type X clade

A) 53 isolates genotyped using 5 unlinked markers identified 9 distinct *T. gondii* genotypes infecting southern sea otters. Type I, II, and III alleles are represented by red, green, and blue, respectively. Sea otters were infected with *T. gondii* strains that possessed either a non-archetypal allele (purple or orange) or a Type II allele (green) at the loci examined. The number of sea otter infections (#SO) and percentage of total infections (%SO)

1187 is shown for each genotype group. Each group was comprised of the following isolates: **A**
1188 (3097, 3142, 3168, 3265, 3483, 3488, 3520, 3637, 3659, 3744, 3786, 3821, 3865, 3947, 3950,
1189 4003, 4045, 4071, 4151), **B** (3458, 3523, 3728, 3897), **C** (3026, 3045, 3077, 3160, 4167), **D**
1190 (3133, 3178, 3183, 3451, 4166), **E** (3819), **F** (3387, 3429, 3503), **G** (3675), **H** (3671), **II**
1191 (2987, 2994, 3005, 3009, 3087, 3131, 3208, 3396, 3521, 3576, 3587, 3636, 3739, 4181).
1192 B) Total Sea Otter Strandings: number of sea otter strandings between 1998-2004 are
1193 charted according to their 5 marker type designation of Type II (green) or sub-types of Type
1194 X (range of purples). Type II infections not only clustered geographically (not shown) but
1195 also temporally while Type X infections are steady over time. Additionally, the Type X, sub-
1196 type A infections which were predominantly avirulent in sea otters show increased sea otter
1197 infections over time.
1198 C) CD1 mice were infected with 50 tachyzoites intraperitoneally with one of 18 isolates
1199 representing all singletons (**E, G, H**) and at least two isolates from each of the other 6 distinct
1200 genotypes (**II, A, B, C, D, F**) recovered from sea otters. Mouse seroconversion and survival
1201 was monitored for 30 days. At least two independent infection experiments were performed
1202 using 5 mice each for every strain tested. Results shown are for those mice that
1203 seroconverted or died acutely during infection. Strains are grouped and colored based on
1204 their genotype and virulence in mice. **Red**, mouse virulent, **Blue**, intermediate mouse
1205 virulent (some mice survived acute infection); **Green**, mouse avirulent (all mice survived
1206 acute infection). Number of mice infected by each genotype and their relative virulence is
1207 indicated.
1208 D) eBURST analysis to identify linkage disequilibrium across 17 nuclear-encoded linked
1209 and unlinked markers identified 4 clonal complexes and 3 unique Type X genotypes among
1210 21 isolates selected from the 9 distinct clades in Figure 1A. Isolate colors represent the
1211 lineage Type from A: red (Type I), green (Type II), blue (Type III), and purple (Type X; all
1212 subgroups (A-H)). Dot size is proportional to the number of isolates with that genotype;
1213 larger dots are multi-isolate genotypes. Clonal complexes (CC) are indicated by lines
1214 connecting isolates and are highlighted with ovals corresponding to their allelic identity (X-
1215 purple, II-green)
1216
1217 **<u>Figure 2</u>: Phylogenetic analyses support the designation of Type X as a recombinant**
1218 **clade of related strains**
1219 Maximum likelihood trees of sequenced markers are depicted. Allele designations were
1220 dependent on a bootstrap support greater than 60%. Isolates are colored based on their
1221 genotype from Figure 1A (I: red, II: green, III: blue, X: purple). A) Mitochondrial marker, CO1.
1222 B) Nuclear genome marker BSR4 on Chromosome IV. At BSR4, all Type X strains possessed
1223 a single allele, referred to as γ that was readily distinguished from Type I, II and III alleles C)
1224 Comparison of two unlinked nuclear-encoded markers GRA7 and BAG1. At both markers,
1225 Type X strains possessed one of two distinct alleles, referred to as γ or δ. Isolates possessing
1226 δ alleles at both markers were highlighted in purple, those that possessed Type II alleles at
1227 both markers were highlighted in green. Isolates that were recombinant and possessed a γ
1228 allele at GRA7 but a δ allele at BAG1 were highlighted in pink, and conversely, a δ allele at
1229 GRA7 and a γ allele at BAG1 were highlighted in orange. D) Comparison of linked genomic
1230 markers GRA7 and SAG4. Genetic recombination is highlighted within this chromosome.
1231 Isolates that are linked and have Type II lineage alleles at both markers are highlighted in

1232 green. Isolates that show recombination between the γ/δ lineage alleles and the Type II
1233 lineage alleles are highlighted in pink and orange depending on recombination
1234 directionality.
1235
1236 **<u>Figure 3</u>: Genome-wide SNP typing of Type X displays haploblock recombination**
1237 **across the genome**
1238 A) CGH-array hybridization of Type X strains against Type I, II and III lineage-specific probes
1239 identified 8 distinct Type II and non-Type II hybridization patterns across the 8 Type X
1240 strains examined. SNPs are represented by a dot in one of three rows indicating where the
1241 isolate has hybridized to the microarray with a hybridization characteristic of lineage Type
1242 I (red), II (green), or III (blue). Grey indicates no hybridization at this location. Chromosomes
1243 are represented as alternating grey and white bars.
1244 B) Linkage disequilibrium and network reticulation demonstrates the existence of recombination
1245 across the genomes of Type X. A NeighborNet tree based on whole genome sequence identified
1246 568,592 SNP variant positions across the 19 Type X strains reference mapped to ME49. Type I
1247 (red), II (green), III (blue), and X (purple) strains were annotated in the box beside the group based
1248 on their previous MLST designation. Two clusters of Type X strains were identified. Strains within
1249 the inset were colored based on their murine virulence: virulent (red), intermediate virulent (blue),
1250 and avirulent (green).
1251
1252 **<u>Figure 4</u>: SNP density plots identify distinct and different recombination crossover**
1253 **points between the Type II and γ/δ lineage among the Type X strains examined**
1254 WGS reads for all strains were mapped to the ME49 reference genome. Each row represents
1255 the SNPs of one strain mapped in a sliding 10 kb window across the genome, separated into
1256 chromosomes columns. Vertical bars across the row represent the number of SNPs, the
1257 height of the bar (from y-axis 0-200 bp) represents the amount of divergence from the ME49
1258 genome. Type X strains were grouped based on their MLST designation.
1259
1260 **<u>Figure 5</u>: Whole genome sequencing establishes Type X as a recombinant clade of**
1261 **related strains**
1262 A) PopNet analysis for all strains sequenced at WGS resolution. The genome of each strain is
1263 represented by a circle of concatenated chromosomes starting at the top with Chromosome
1264 Ia and rotating clockwise to Chromosome XII. All strains were clustered into four distinct
1265 color groups based on their degree of shared ancestry (Type II-green, Type X-pink, Type I-
1266 cyan, Type III-blue) and each isolate has an inner circle that is arrayed in color-group
1267 haploblocks based on shared ancestry. Line thickness between strains indicates their
1268 interrelatedness, boldest lines indicate 90-100% of the genome is in linkage disequilibrium,
1269 as shown for strains 4167, 3819, 3045 and 3160.
1270 B) Genetic model showing cross-over recombination points for isolate 3142. Type II ancestry
1271 is shown in green whereas the mosaic of γ/δ ancestry is shown in black.
1272 C) 21 WGS isolates are shown with their group designation for the 5 loci MLST (Figure 1A),
1273 eBURST clonal complex identity (Figure 1D), 17 loci MLST (Table 1), and murine virulence
1274 phenotype (Figure 1C).
1275
1276 **<u>Figure 6:</u> Population level QTL analysis identifies multiple chromosomal locations**
1277 **associated with acute murine virulence**

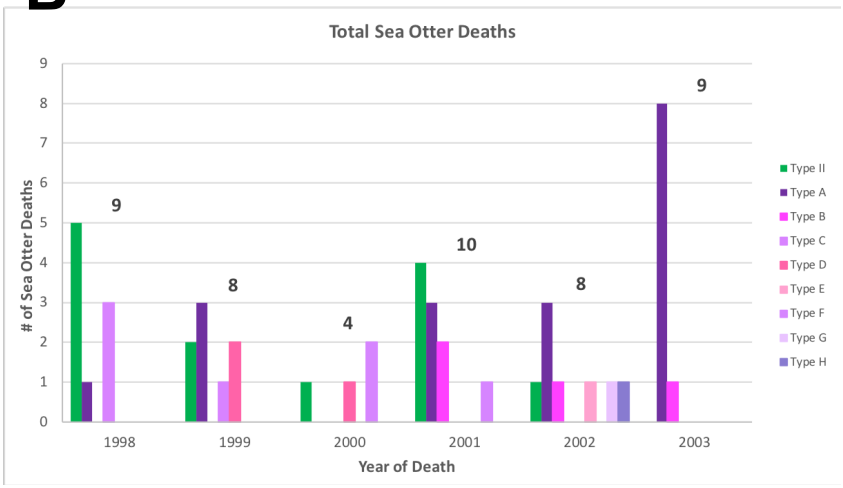
1278 A) WGS data from Type X strains was down-selected to one SNP every 5 kb and analyzed
1279 using standard QTL methods with 1000 bootstrap support. LOD scores are shown across the
1280 chromosomes for these strains based on genetic association with acute murine virulence as
1281 shown in Table 1.
1282 B) Significantly associated peaks from the QTL were identified based on clusters of SNP
1283 peaks. The LOD scores for the tallest portion of the associated genomic region are listed.
1284 C) Distance-based Neighbor-joining ROP33 tree. Alignment of *ROP33* DNA sequences were
1285 used to construct a phylogenetic tree representing the majority of ROP33 alleles encoded
1286 within the species *Toxoplasma gondii.* Mouse virulence for each isolate is depicted by a color-
1287 code, per Figure 1C. **Red**, mouse virulent, **Blue**, intermediate mouse virulent (some CD-1
1288 mice survived acute infection); **Green**, mouse avirulent (all CD-1 mice survived acute
1289 infection). Inoculum size was 50 tachyzoites injected intraperitoneally in CD-1 outbred mice.
1290 D) Type I RH parasites deficient in ROP18 and ROP33 are less virulent than Type I RH
1291 parasites deficient in ROP18 alone, indicating that ROP33 is a virulence locus in *Toxoplasma.*
1292 CD-1 mice were infected intraperitoneally with 500 tachyzoites of the RH*Δrop18* or
1293 RH*Δrop18Δrop33* DKO, only the results for seropositive mice are shown. Data are combined
1294 from two independent experiments, each with five mice per group.
1295 E) Bioluminescent detection of parasite burden *in vivo* at 9 days post-infection of CD-1 mice
1296 with 50 tachyzoites injected intraperitoneally. A representative image with photon output
1297 in photons/second/cm2/surface radiance (*sr*) is shown.
1298
1299 **<u>Supplemental Figure 1</u>: Stranding location, necropsy date, sex, *Toxoplasma* titer**
1300 **(IFAT), and co-infection status of every California Sea otter that yielded a *Toxoplasma***
1301 **isolate for genetic characterization**
1302 A) Geographical map for all *Toxoplasma*-infected sea otter stranding locations along the
1303 California coastline. Each Sea otter stranding location is represented by a black dot.
1304 B) Isolates of *Toxoplasma gondii* used in this paper. Isolates in yellow were genotyped
1305 previously using PCR-RFLP analyses at 4 loci (*B1, SAG1, SAG2, SAG3*) and isolates 3131, 3133,
1306 3160, 3265 were DNA sequenced at the *GRA6 locus* [25]. ATOS numbers and stranding
1307 locations identify the location of each stranded Sea otter. Isolates were grouped together
1308 based on their 5 locus DNA sequence genotype and then by their ATOS number. "Tg"
1309 indicates *Toxoplasma gondii,* "Sn" indicates co-infection with *Sarcocystis neurona.*
1310
1311 **<u>Supplemental Table 1</u>: Expanded sequencing marker genotyping of Type X isolates**
1312 **reveals chromosomal segregation and recombination within chromosomes**
1313 21 isolates characterized at 17 nuclear markers (ROP1 was excluded as it is a microsatellite
1314 marker, and prone to elevated mutation rates) clade into 12 distinct Type X genotypes (X1-
1315 X12). Isolate numbers in red were sequenced at whole-genome (WGS) resolution. Types I,
1316 II, III, γ and δ lineage alleles, as determined by phylogenetic comparisons, are colored red,
1317 green, blue, purple, and orange, respectively. Shades of colors represent genetic drift below
1318 the 60% bootstrap delineation from the canonical allele. White represents uninformative or
1319 incomplete DNA sequencing results.
1320
1321 **<u>Supplemental Table 2</u>: Potential virulence candidate genes derived from the QTL**
1322 **analysis**

1323 Chromosome peaks with LOD scores over 3 as shown in Figure 3B were interrogated via
1324 ToxoDB to select genes in these regions which have predicted signal sequences (Signal) and
1325 transmembrane (TM) domains in *Toxoplasma.* Genes are listed with gene ID, genomic
1326 location, and predicted protein function.

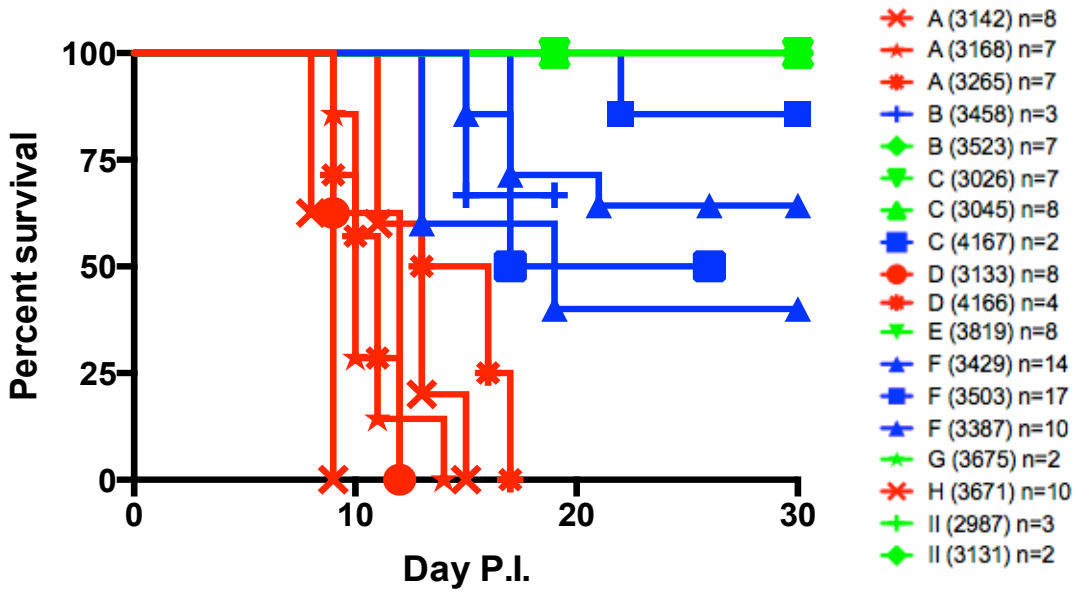
A

Marker	BSR4	BAG1	GRA6	ROP1	SAG3	# SO	% SO
Marker Type	Seq	Seq	Seq	RFLP	Seq		
Chromosome	IV	VIIb	X	XI	XII		
Length	1138	1917	758	Dde1	253		
Type I							
Type II							
Type III							
II						14	26
A	32	44	44			19	36
B	32	44	32	32		4	7.5
C	32	32	32			5	9.4
D	32	44	44			5	9.4
E	32	32	44			1	1.9
F	32	44	32			3	5.7
G				44		1	1.9
H	32	44				1	1.9

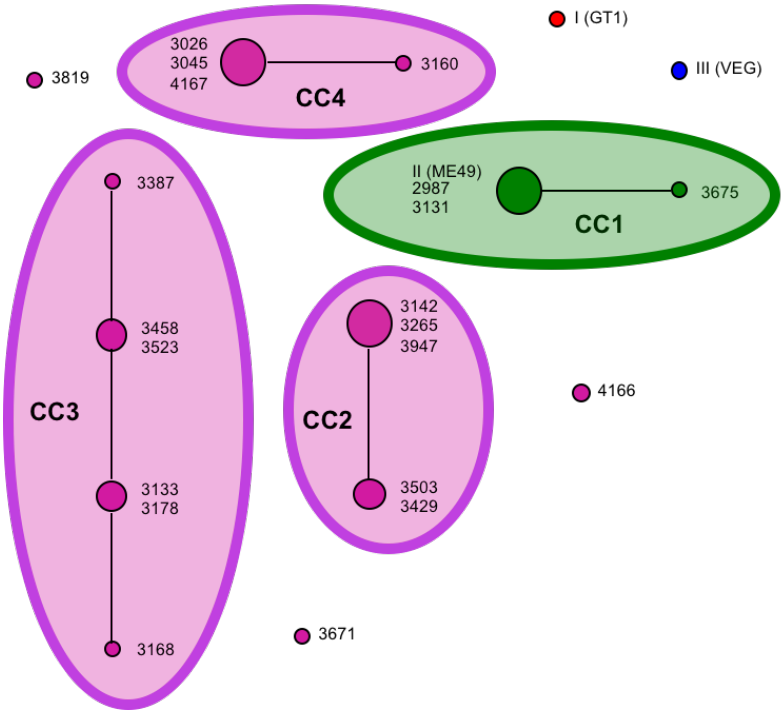
B

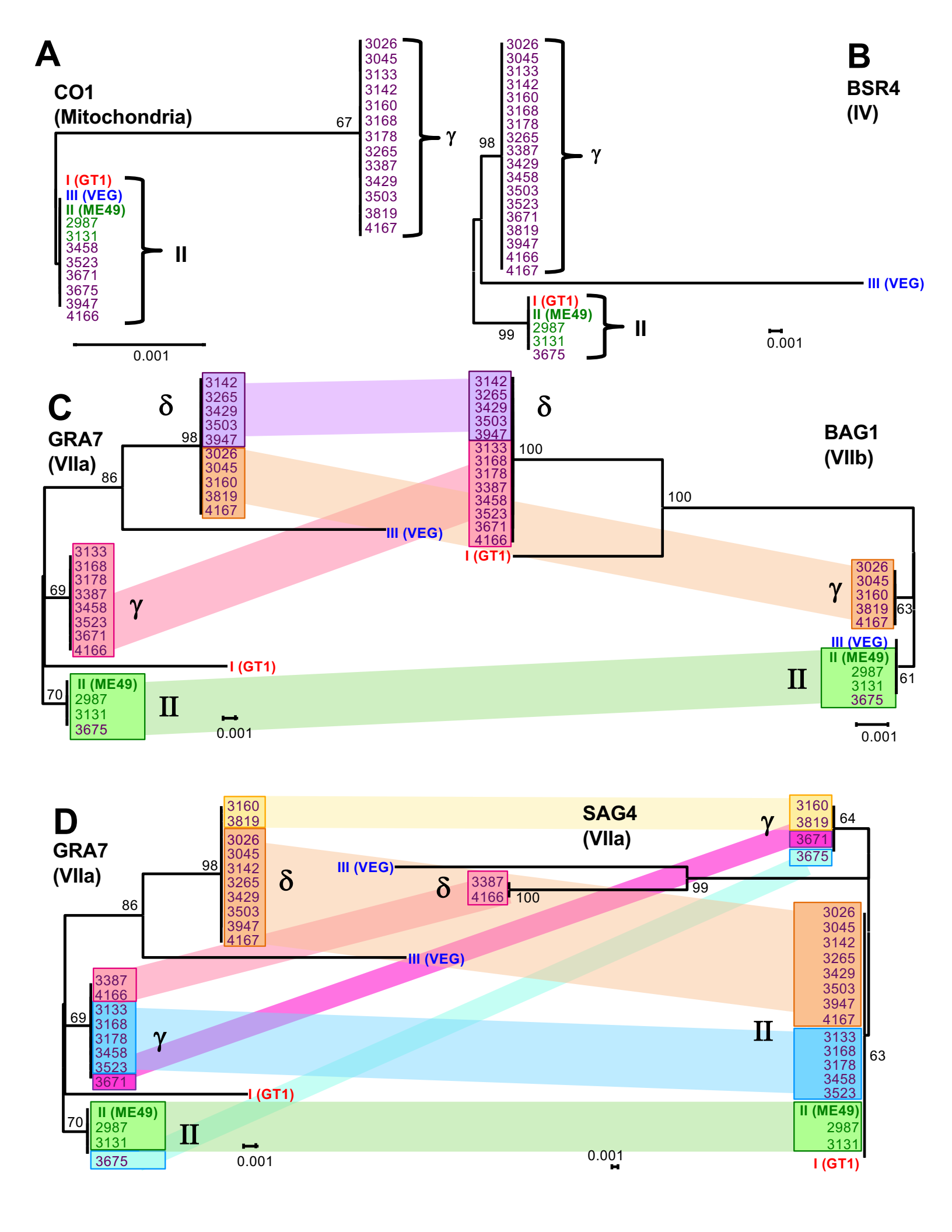


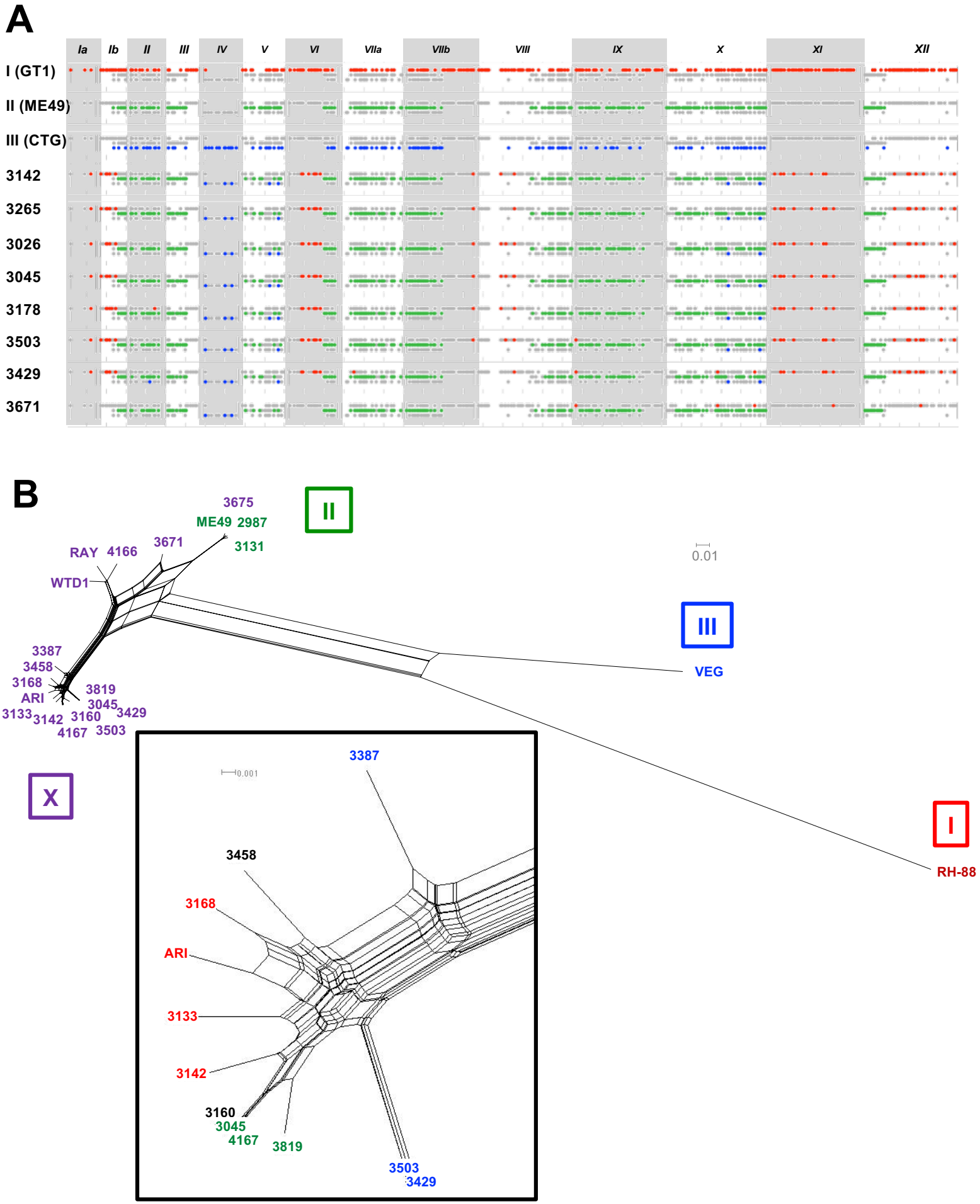
C



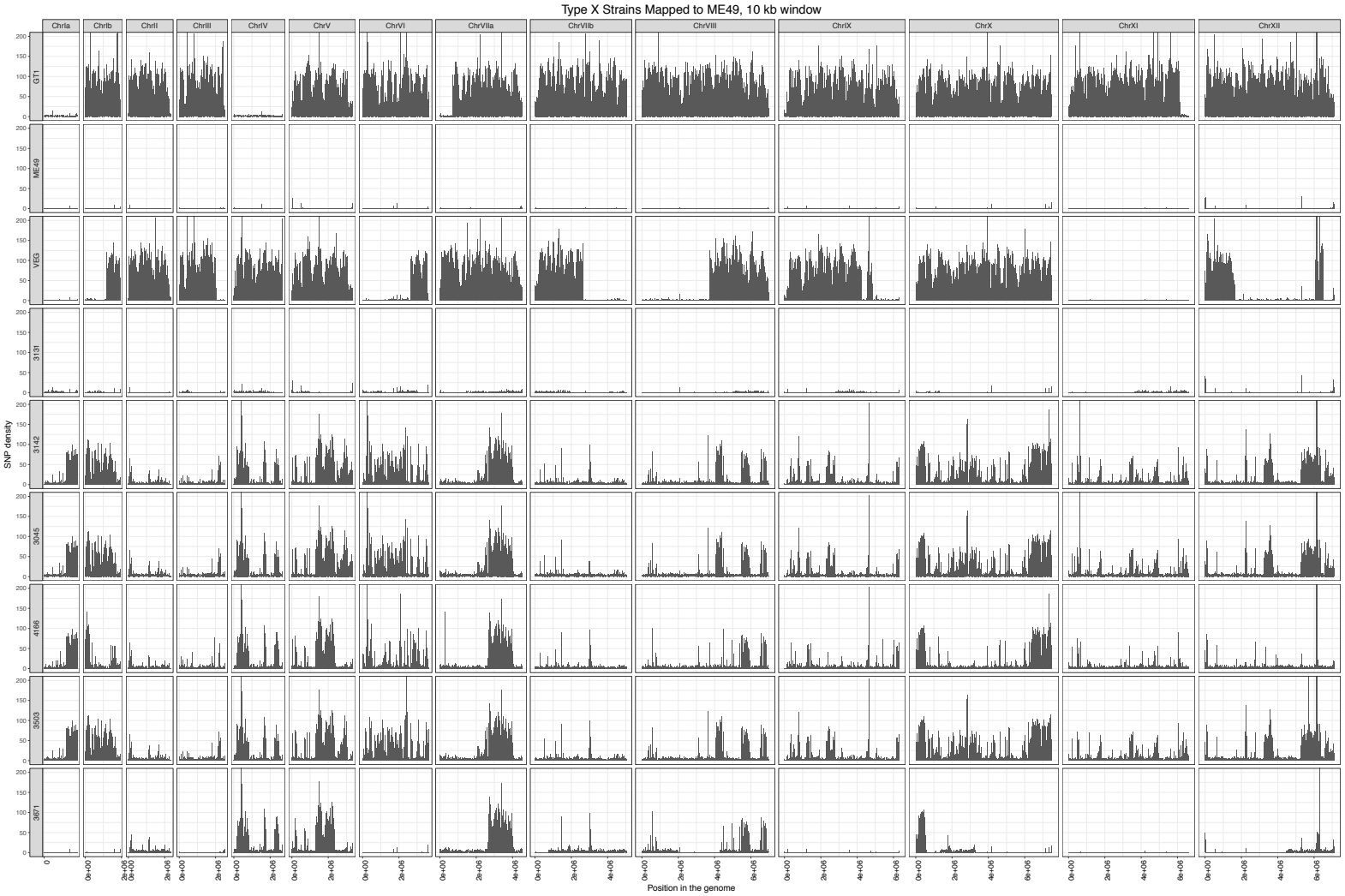
D



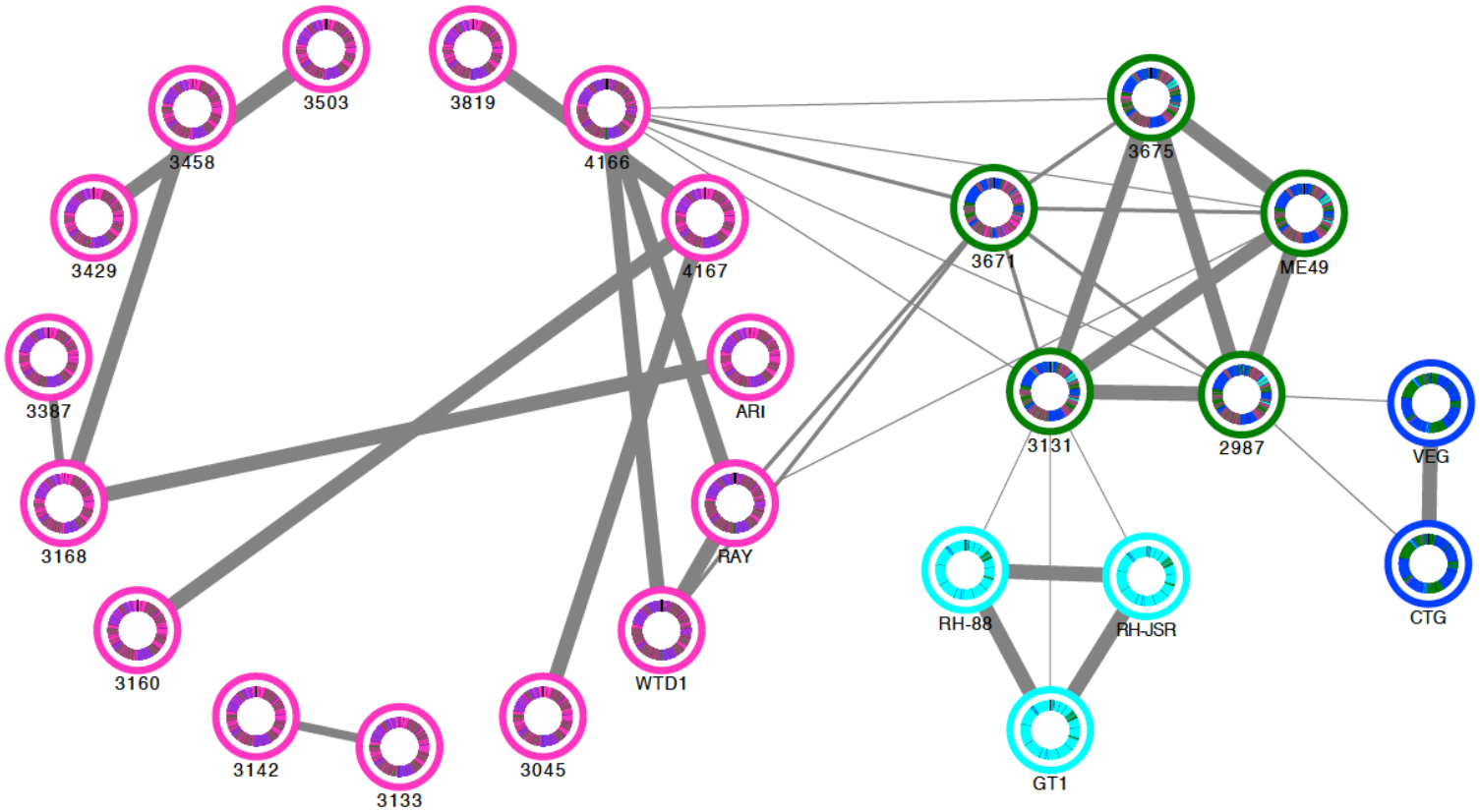

A
CO1
(Mitochondria)
67
3026
3045
3133
3142
3160
3168
3178
3265
3387
3429
3503
3819
4167
γ
I (GT1)
III (VEG)
II (ME49)
2987
3131
3458
3523
3671
3675
3947
4166
II
0.001
B
BSR4
(IV)
98
3026
3045
3133
3142
3160
3168
3178
3265
3387
3429
3458
3503
3523
3671
3819
3947
4166
4167
γ
III (VEG)
I (GT1)
II (ME49)
2987
3131
3675
99
II
0.001
C
GRA7
(VIIa)
δ
3142
3265
3429
3503
3947
98
3026
3045
3160
3819
4167
86
III (VEG)
3133
3168
3178
3387
3458
3523
3671
4166
69
γ
I (GT1)
II (ME49)
2987
3131
3675
70
II
0.001
δ
3142
3265
3429
3503
3947
100
3133
3168
3178
3387
3458
3523
3671
4166
I (GT1)
100
BAG1
(VIIb)
γ
3026
3045
3160
3819
4167
63
III (VEG)
II (ME49)
2987
3131
3675
61
II
0.001
D
GRA7
(VIIa)
3160
3819
98
3026
3045
3142
3265
3429
3503
3947
4167
δ
86
III (VEG)
3387
4166
69
3133
3168
3178
3458
3523
3671
γ
I (GT1)
II (ME49)
2987
3131
3675
70
II
0.001
SAG4
(VIIa)
δ
3387
4166
100
III (VEG)
99
γ
3160
3819
64
3671
3675
3026
3045
3142
3265
3429
3503
3947
4167
II
3133
3168
3178
3458
3523
63
II (ME49)
2987
3131
I (GT1)
0.001


A
Ia
Ib
II
III
IV
V
VI
VIIa
VIIb
VIII
IX
X
XI
XII
I (GT1)
II (ME49)
III (CTG)
3142
3265
3026
3045
3178
3503
3429
3671
B
3675
ME49 2987
3131
II
RAY 4166
3671
WTD1
0.01
III
VEG
3387
3458
3168
ARI
3133
3142
3160
3819
3045
3429
4167
3503
X
I
RH-88
0.001
3387
3458
3168
ARI
3133
3142
3160
3045
4167
3819
3503
3429

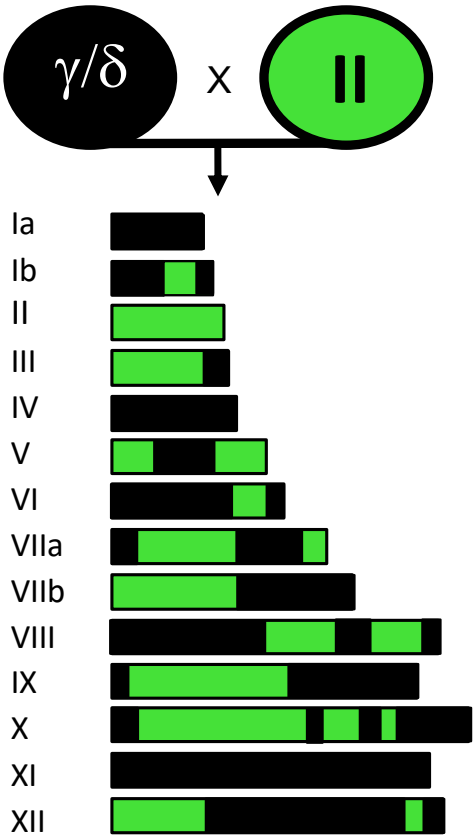
A


Type X Strains Mapped to ME49, 10 kb window
ChrIa
ChrIb
ChrII
ChrIII
ChrIV
ChrV
ChrVI
ChrVIIa
ChrVIIb
ChrVIII
ChrIX
ChrX
ChrXI
ChrXII
GT1
ME49
VEG
3131
3142
3045
4166
3503
3671
SNP density
Position in the genome

A



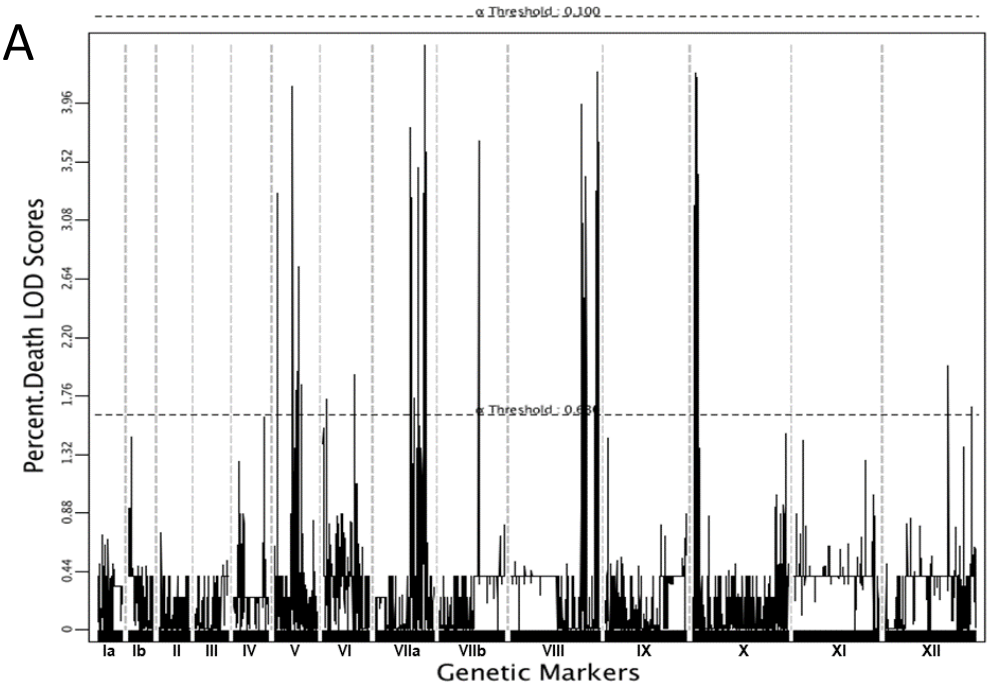
B



C

Isolate	5 loci	CC	17 loci	Mouse VIR
2987	II	CC1	II	-
3131	II	CC1	II	-
3142	A	CC2	X1	+
3265	A	CC2	X1	+
3947	A	CC2	X1	n.d.
3168	A	CC3	X2	+
3458	B	CC3	X3	+/-
3523	B	CC3	X3	-
3026	C	CC4	X4	-
3045	C	CC4	X4	-
4167	C	CC4	X4	+/-
3160	C	CC4	X5	n.d.
3133	D	CC3	X6	+
3178	D	CC3	X6	n.d.
4166	D	-	X7	+
3819	E	-	X8	-
3503	F	CC2	X9	+/-
3429	F	CC2	X9	+/-
3387	F	CC3	X10	+/-
3675	G	CC1	X11	-
3671	H	-	X12	+

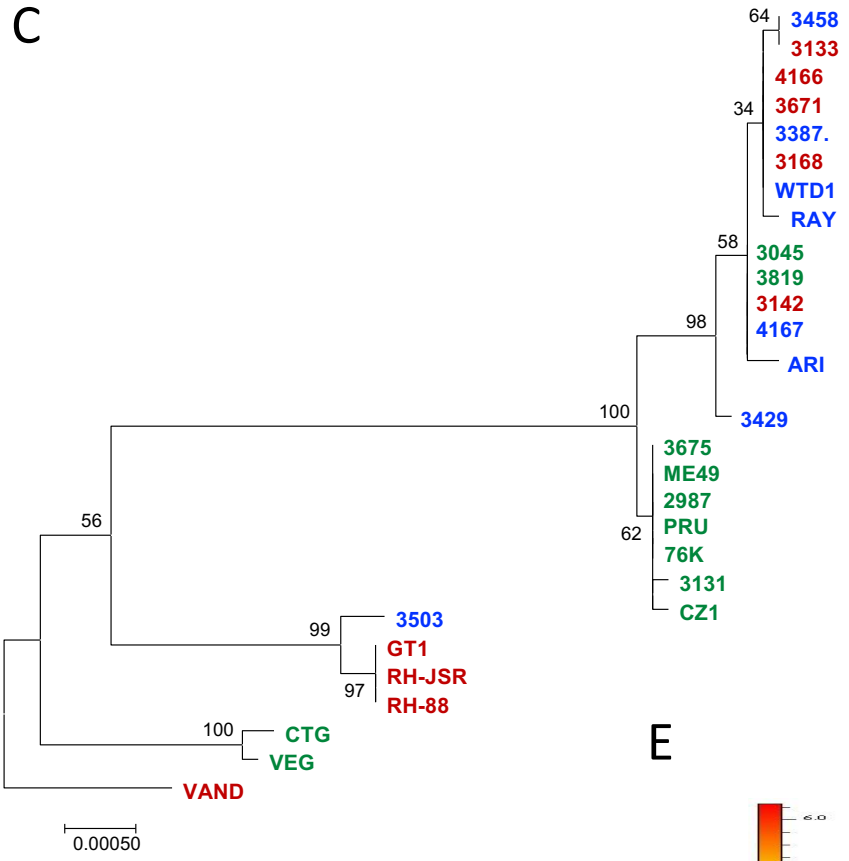
A



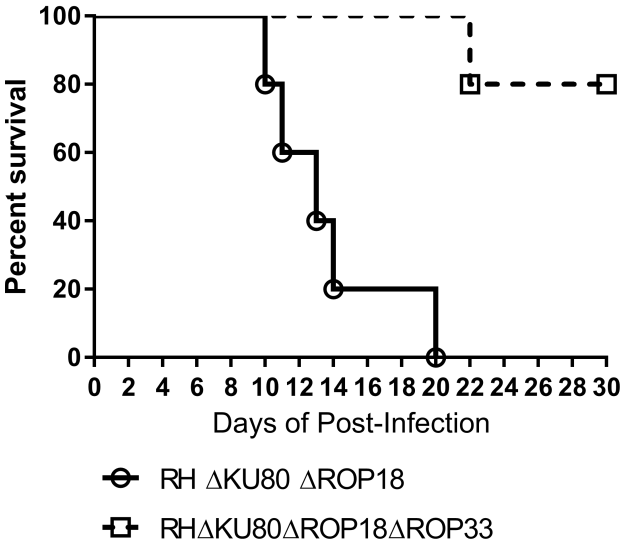
B

Chrm	Start (bp)	End (bp)	LOD
V	1430036	2165042	4.091
VIIa	3815011	4085005	4.400
VIIb	3035265	3160030	3.680
VIII	6745008	6905157	4.194
X	7314	290066	4.190

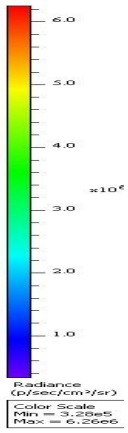
C

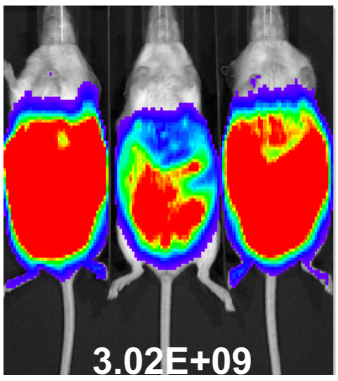


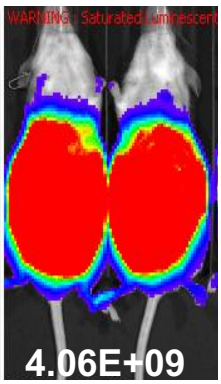
D



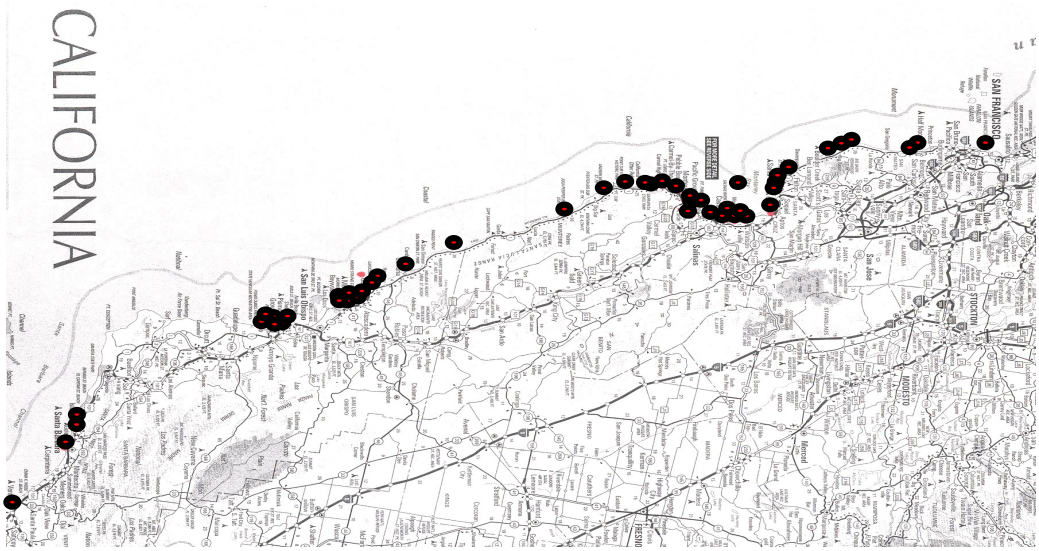
E







A



B

Isolate	ATOS #	5 loci	17 loci	Stranding Location	Necropsy Date	Sex	IFAT	Infection
2994	249	II		Santa Cruz	05/15/98	M	no serum	Tg
3739	264	II		Santa Cruz 26th Ave Beach	06/12/02	M	2560	Tg
3576	274	II		New Brighton State Beach	09/05/01	M	20480	Tg
3587	281	II		Rio Del Mar Beach	09/22/01	M	1280	Tg
2987	291	II	II	Mannressa State Beach	05/11/98	F	1280	Tg
4181	303	II		Sunset State Beach	04/14/04	F	40960	Sn/Tg
3009	321	II		Elkhorn Slough	06/04/98	M	640	Tg
3521	325	II		Salinas River State Beach	05/19/01	M	2560	Tg
3208	357	II		Monterey Bay	07/12/99	M	5120	Tg
3636	376	II		Del Monte Beach	12/27/01	M	80	Tg
3087	377	II		Monterey Harbor	10/28/98	F	640	Tg
3131	379	II	II	Monterey Bay	02/05/99	M	320	Tg
3005	928	II		Pismo Beach	06/01/98	M	160	Tg
3396	928	II		Oceano Dunes	11/11/00	M	20480	Tg
3097	84	A		Dunes Beach, San Mateo	11/16/98	M	5120	Tg
3142	180	A	X1	Ano Nuevo Island	02/24/99	F	320	Tg
3786	180	A		Ano Nuevo Island	09/20/02	M	2560	Sn
4045	260	A		Seabright Beach	10/21/03	M	10240	Tg
3950	263	A		Santa Cruz, 21st Street Beach	07/16/03	M	10240	Sn/Tg
3520	306	A		Sunset Beach	05/16/01	M	320	Tg
3659	374	A		Del Monte Beach	02/28/02	F	40	Tg
3265	426	A	X1	Carmel Beach	09/28/99	F	10240	Tg
3168	450	A	X2	Monterey Bay	04/09/99	F	2560	Tg
3865	480	A		Carmel River State Beach	04/03/03	F	10240	Tg
3947	523	A	X1	Pfieffer Burns Beach	06/26/03	F	20480	Tg
4003	750	A		Cambria	09/17/03	F	10240	Tg
4071	807	A		Cayucos, North of E Steet	12/03/03	F	2560	Tg
3483	820	A		Morro Strand	03/30/01	F	320	Sn/Tg
4151	825	A		Morro Bay	04/04/04	F	>40960	Sn/Tg
3488	919	A		Pismo Beach, ~400m North of Grand Ave	04/04/01	F	640	Tg
3821	922	A		Pismo Beach	01/06/03	M	2560	Tg
3744	924	A		Pismo Beach	07/02/02	F	5120	Tg
3637	926	A		Oceana Dunes	02/14/03	F	no serum	Tg
3523	141	B	X3	Pescadero Point	05/31/01	F	5120	Sn/Tg
3458	422	B	X3	Stillwater Cove	01/16/01	F	<80	Tg
3897	435	B		Pt Lobos State Reserve	04/22/03	F	10240	Tg
3728	1,445	B		Ventura County (Pt Mugu St Pk)	05/29/02	M	320	Tg
3077	145	C		Bean Hollow, Half Moon Bay	09/27/98	M	1280	Tg
3045	808	C	X4	Cayucas	07/09/98	F	no serum	Tg
3026	840	C	X4	Morro Bay Sandspit	06/24/98	M	320	Tg
4167	852	C	X4	Morro Creek	04/09/04	M	10240	Sn/Tg
3160	933	C	X5	Pismo Beach	03/29/99	M	320	Tg
3178	321	D	X6	Moss Landing Harbor	03/26/02	F	320	Tg
3183	816	D		Morro Strand	05/04/99	F	640	Tg
3133	818	D	X6	Morro Strand Campground	02/04/99	M	20480	Tg
3451	836	D		Morro Bay, Sandspit between port 3 & 4	11/29/00	M	<80	Tg
4166	827	D	X7	Morro Bay	04/09/04	M	40960	Sn/Tg
3819	732	E	X8	San Simeon Bay	12/28/02	F	5120	Tg
3429	827	F	X9	Morro Bay	10/30/00	F	2560	Tg
3387	922	F	X10	Pismo Beach	07/19/00	F	1280	Tg
3503	1,319	F	X9	Rincon Pt., Ventura County	08/09/01	M	2560	Tg
3675	435	G	X11	Point Lobos	03/20/02	F	1280	Tg
3671	321	H	X12	Elkhorn Slough	03/13/02	M	640	Sn/Tg
		n=53	n=21					

Marker	5 Loci	17 Loci	C228	Chr_II_10	C292	BSR4	L358	PK1	SAG4	ROP18	GRA7	BAG1	SRS2	SAG1	SAG2	BTB	GRA6	ROP1	UPRT	SAG3	APICO	COX1
Chromosome			Ib	II	III	IV	V	VI	VIIa	VIIa	VIIa	VIIb	VIII	VIII	VIII	IX	X	XI	XI	XII	API	MITO
Length			421	639	397	1138	354	1011	174	1617	460	1917	983	700	1243	331	758	960	1214	253	442	418
Type I		I	α	α	α	Type II	α	α	Type II	α	α	α	α	α	α	α	α	α	α	α	α	Type II
Type II		II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II
Type III		III	β	β	β	β	β	β	β	β	β	Type II	Type II	Type II	β	β	β	Type II	Type II	β	β	Type II
2987	II	II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	α	Type II
3131	II	II	Type II	Type II	Type II	Type II	Type II	γ	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II
3142	A	X1	Type II	Type II	Type II	γ	α	δ	Type II	γ	δ	δ	γ	γ	Type II	γ	δ	γ	Type II	Type II	α	γ
3265	A	X1	Type II		Type II	γ	α	δ	Type II	γ	δ	δ	γ	γ	Type II	γ	δ	γ		Type II	α	γ
3947	A	X1	Type II			γ	α	δ	Type II	γ	δ	δ	γ	γ	Type II	γ	δ	δ		Type II	α	
3168	A	X2	Type II	Type II	Type II	γ	α	Type II	Type II	γ	γ	δ	γ	γ	Type II	γ	δ	δ	Type II	Type II	α	γ
3458	B	X3	Type II	Type II	Type II	γ	α	γ	Type II	γ	γ	δ	γ	γ	Type II	γ	γ		Type II	Type II	γ	Type II
3523	B	X3	Type II		Type II	γ	α	γ	Type II	γ	γ	δ	γ	γ	Type II	γ	γ		Type II	Type II	γ	Type II
4167	C	X4	Type II	Type II	Type II	γ	α	δ	Type II	δ	δ	γ	γ	γ	Type II	γ	γ	γ	Type II	Type II	α	γ
3026	C	X4	Type II	Type II	Type II	γ	α	δ	Type II	δ	δ	γ	γ	γ	Type II	γ	γ	γ	Type II	Type II	α	γ
3045	C	X4	Type II	Type II	Type II	γ	α	δ	Type II	δ	δ	γ	γ	γ	Type II	γ	γ	γ	Type II	Type II	α	γ
3160	C	X5	Type II	Type II	Type II	γ	α	δ	γ	δ	δ	γ	γ	γ	Type II	γ	γ	γ	Type II	Type II	α	γ
3133	D	X6	Type II	Type II	Type II	γ	α	γ	Type II	γ	γ	δ	γ	γ	Type II	γ	δ		Type II	Type II	α	γ
3178	D	X6	Type II	Type II	Type II	γ	α	γ	Type II	γ	γ	δ	γ	γ	Type II	γ	δ	Type II	Type II	Type II	α	γ
4166	D	X7	Type II	Type II	Type II	γ	α	δ	γ	γ	γ	δ	γ	γ	Type II	γ	δ		Type II	Type II	α	Type II
3819	E	X8	Type II	Type II	Type II	γ	α	δ	γ	γ	δ	γ	γ	γ	Type II	γ	δ	Type II	Type II	Type II	α	γ
3503	F	X9	Type II	Type II	Type II	γ	α	δ	Type II	γ	δ	δ	γ	γ	Type II	γ	γ	γ	Type II	Type II	α	γ
3429	F	X9	Type II	Type II	Type II	γ	α	δ	Type II	γ	δ	δ	γ	γ	Type II	γ	γ	γ	Type II	Type II	Type II	γ
3387	F	X10	Type II	Type II	Type II	γ	α	γ	γ	γ	γ	δ	γ	γ	Type II	γ	γ	δ	Type II	Type II	α	γ
3675	G	X11	Type II	Type II	Type II	Type II	Type II	Type II	γ	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II		Type II	Type II	α	Type II
3671	H	X12	Type II	Type II	Type II	γ	α	Type II	γ	γ	γ	δ	Type II	Type II	Type II	Type II	Type II	Type II	Type II	Type II	α	Type II

Type II	Type II
α	α
β	β
γ	γ
δ	δ

Chr	Start (bp)	End (bp)	LOD	TM	Signal	Gene ID	Genomic Location(s)	Product Description
5	220001	350021	3.25	Y	Y	TGME49_220240	TGME49_chrV:235,019..240,463(-)	hypothetical protein
				Y	Y	TGME49_220300	TGME49_chrV:275,453..283,005(-)	ribosomal protein L15 protein
				Y	Y	TGME49_220330	TGME49_chrV:295,153..299,369(-)	hypothetical protein
				Y	Y	TGME49_220350	TGME49_chrV:310,715..318,211(+)	tRNA ligases class II (D, K and N) domain-containing protein
5	1430036	2165042	4.09	Y	Y	TGME49_285870	TGME49_chrV:2,158,544..2,162,541(-)	SAG-related sequence SRS20A
				Y	Y	TGME49_285940	TGME49_chrV:2,123,872..2,135,543(-)	hypothetical protein
				Y	Y	TGME49_286180	TGME49_chrV:1,933,873..1,942,288(-)	tRNA ligases class I (M) protein
				Y	Y	TGME49_286450	TGME49_chrV:1,824,587..1,826,598(+)	dense granule protein GRA5
				Y	Y	TGME49_286530	TGME49_chrV:1,790,919..1,793,716(+)	hypothetical protein
				Y	Y	TGME49_286620	TGME49_chrV:1,754,511..1,762,820(-)	S1 RNA binding domain-containing protein
				Y	Y	TGME49_286630	TGME49_chrV:1,750,888..1,752,970(+)	redoxin domain-containing protein
				Y	Y	TGME49_286770	TGME49_chrV:1,682,413..1,683,569(+)	hypothetical protein
				Y	Y	TGME49_213920	TGME49_chrV:1,428,849..1,440,483(-)	hypothetical protein
7a	3815011	4085005	4.40	Y	Y	TGME49_201130	TGME49_chrVIIa:4,019,644..4,026,659(+)	rhoptry kinase family protein ROP33
				Y	Y	TGME49_201180	TGME49_chrVIIa:3,982,746..3,988,716(-)	hypothetical protein
				Y	Y	TGME49_201390	TGME49_chrVIIa:3,908,769..3,911,287(-)	hypothetical protein
				Y	Y	TGME49_201750	TGME49_chrVIIa: 3,822,835 – 3,823,206(+)	Tctex-1 family protein
				Y	Y	TGME49_201760	TGME49_chrVIIa: 3,816,484 – 3,820,369(+)	hypothetical protein
7b	3035265	3160030	3.68	Y	Y	TGME49_258828	TGME49_chrVIIb:3,159,970..3,160,907(+)	hypothetical protein
				Y	Y	TGME49_258840	TGME49_chrVIIb:3,137,682..3,141,698(+)	hypothetical protein
				Y	Y	TGME49_258860	TGME49_chrVIIb:3,122,191..3,126,503(+)	hypothetical protein
				Y	Y	TGME49_258890	TGME49_chrVIIb:3,103,220..3,104,003(-)	hypothetical protein
				Y	Y	TGME49_258910	TGME49_chrVIIb:3,092,653..3,098,081(+)	hypothetical protein
				Y	Y	TGME49_258950	TGME49_chrVIIb:3,079,180..3,083,732(+)	lectin family protein
				Y	Y	TGME49_258990	TGME49_chrVIIb:3,038,025..3,049,513(+)	Bromodomain-containing protein
				Y	Y	TGME49_259000	TGME49_chrVIIb:3,035,953..3,049,513(+)	hypothetical protein
8	5435051	6125010	3.8	Y	Y	TGME49_269150	TGME49_chrVIII:6,039,099..6,042,614(-)	DHHC zinc finger domain-containing protein
				Y	Y	TGME49_269190	TGME49_chrVIII:6,002,854..6,007,557(-)	glyceraldehyde-3-phosphate dehydrogenase GAPDH2
				Y	Y	TGME49_269400	TGME49_chrVIII:5,836,645..5,840,235(+)	oxidoreductase, short chain dehydrogenase/reductase family protein
				Y	Y	TGME49_269750	TGME49_chrVIII:5,625,999..5,628,567(+)	CrcB family protein
				Y	Y	TGME49_269885	TGME49_chrVIII:5,564,529..5,572,853(+)	rhoptry metalloprotease toxolysin TLN1
				Y	Y	TGME49_269920	TGME49_chrVIII:5,535,689..5,545,827(-)	phosphatidylserine decarboxylase
				Y	Y	TGME49_269980	TGME49_chrVIII:5,483,536..5,489,390(-)	preprotein translocase Sec61, putative
				Y	Y	TGME49_269690	TGME49_chrVIII:5,680,572..5,684,829(+)	hypothetical protein
				Y	Y	TGME49_270030	TGME49_chrVIII:5,448,751..5,449,486(+)	hypothetical protein
8	6745008	6905157	4.19	Y	Y	TGME49_200230	TGME49_chrVIII:6,753,611..6,755,422(-)	microneme protein MIC17C
				Y	Y	TGME49_200440	TGME49_chrVIII:6,883,172..6,887,049(+)	hypothetical protein
10	280002	290066	4.19	N/A	N/A	N/A	N/A	N/A