

Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp1_A		Cluster1	1	1
Sp2_A		Cluster1	1	1
Sp3_A		Cluster1	1	1
Sp1_B		Cluster2	1	1
Sp2_B		Cluster2	1	1
Sp3_B		Cluster2	1	1
Sp1_C		Cluster3	1	1
Sp2_C		Cluster3	1	1
Sp3_C		Cluster3	1	1

Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp1_scf4		Cluster4	1	2
Sp1_scf4		Cluster5	1	2

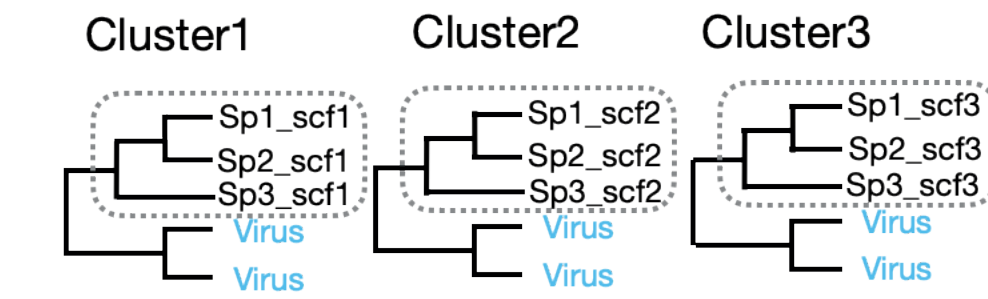
Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp1_scf1		Cluster6	1	3
Sp2_scf2		Cluster6	1	3
Sp1_scf1		Cluster7	1	3
Sp2_scf3		Cluster7	1	3

Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp6_scf1		Cluster8	1	4
Sp10_scf2		Cluster8	2	5
Sp6_scf2		Cluster9	1	4
Sp10_scf3		CLuster9	2	5

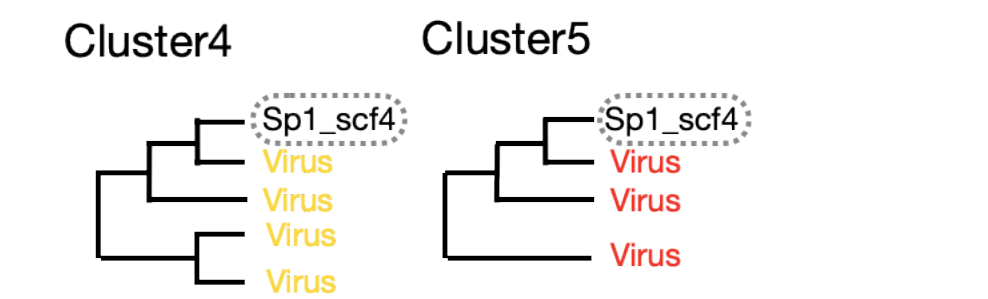
Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp4_scf1		Cluster10	1	6
Sp5_scf2		Cluster10	1	6
Sp4_scf2		Cluster11	1	7
Sp5_scf3		Cluster11	1	7
Sp4_scf3		Cluster12	1	7
Sp5_scf4		Cluster12	1	7

Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp3_scf1		Cluster13	1	8
Sp3_scf2		Cluster13	1	8
Sp3_scf3		Cluster13	1	8
Sp3_scf4		Cluster13	1	8
Sp3_scf5		Cluster14	1	8

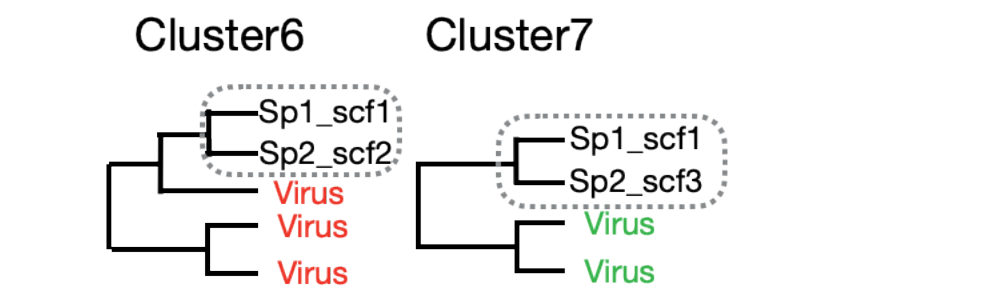
Sp names	Viral Family	Cluster number	Monophyletic clade number	Event number
Sp1_scf1		Cluster15	1	9
Sp1_scf2		Cluster16	1	9
Sp2_scf1		Cluster16	2	10
Sp1_scf3		Cluster17	1	12
Sp2_sc2		Cluster17	2	11



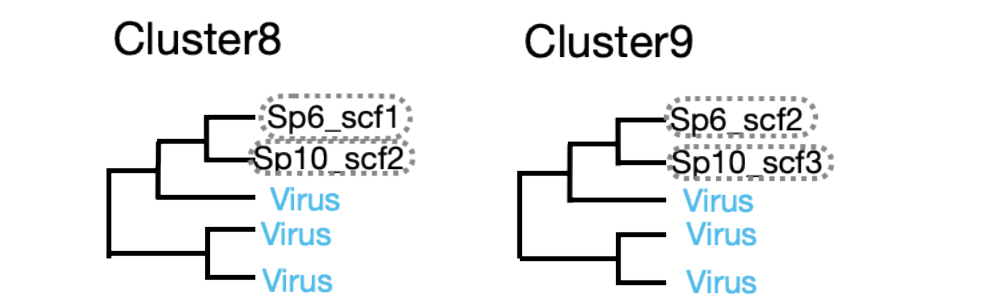
In this example with three clusters (and three distinct genes), three species are involved (SP1,2 and 3). The three species form a monophyletic clade in the species phylogeny of each of the three gene phylogenies and share homology with the same family of blue viruses. As a result, this pattern is consistent with the hypothesis that SP1, 2, and 3' share these genes from their common ancestor. Therefore, we combine these 9 loci into a single event (1) that consists of 3 common EVEs.



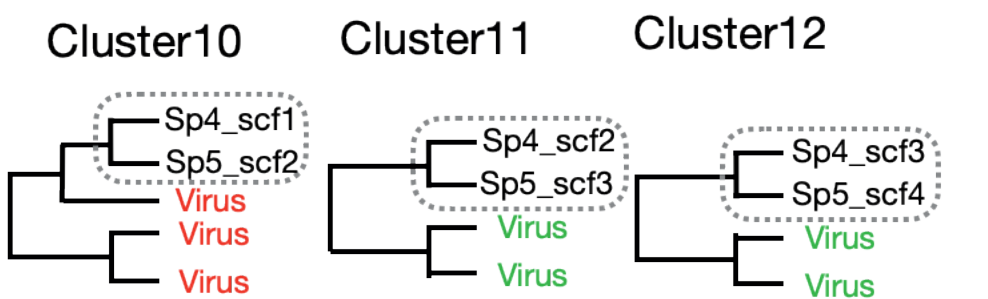
In this example, there are only 1 species and 2 clusters (and consequently 2 different genes) (SP1). We group the SP1 scf4 and SP1 scf4 loci together as a single endogenization event (2) because they share homologies with two distinct viral families (red and yellow), but are both present in the same scaffold (\_scf4). This is done because two independent endogenizations are unlikely to occur at the same place in the genome (note, we restrict this rule to viral families belonging to the same genomic structure).



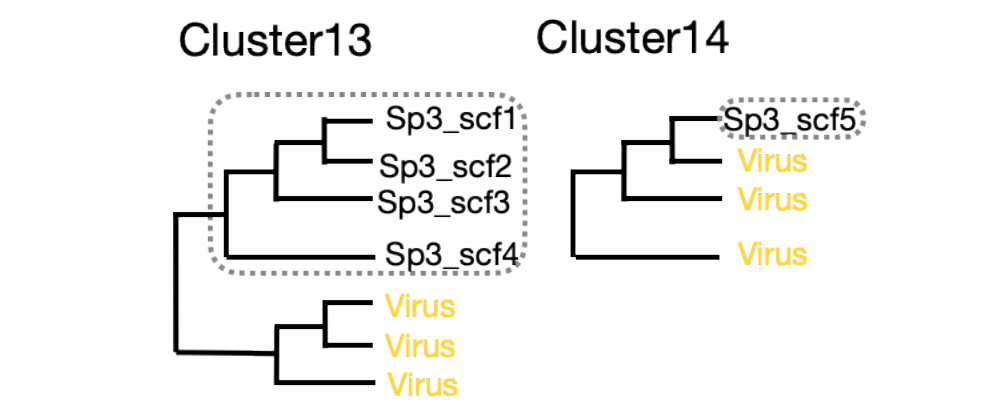
Two species are involved in this example (SP1 and 2), which has two clusters (and thus two different genes). In terms of species phylogeny, the two species are represented by a monophyletic clade in both gene phylogenies. The two loci in SP1 are on the same scaffold (\_scf1), so we will assume that these two genes were endogenized together in the common ancestor of SP1 and SP2, despite the fact that they show homology to a different viral family (note, we restrict this rule to viral families belonging to the same genomic structure). Therefore, by extension, we combine SP2\_scf3 and SP2\_scf2 in the same endogenization event using the information we gained about SP1 (possibly as a result of better assembly) (3).



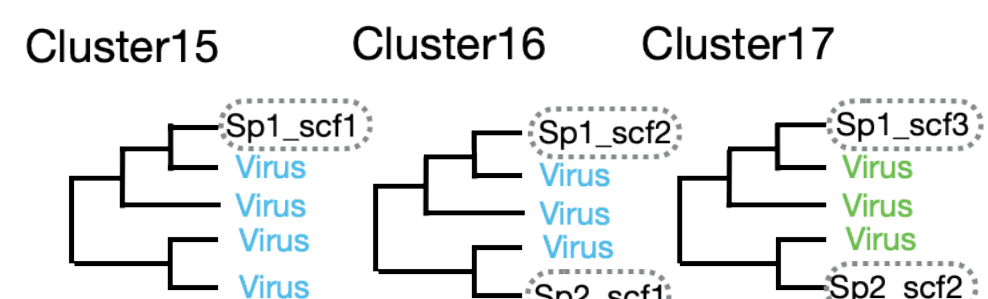
Two species are involved in this example (SP6 and SP10), which has two clusters (and thus two different genes). We treat these two endogenizations as independent in these two species because they do not form a monophyletic clade in terms of species phylogeny (too many species within the same clade lack the EVEs). Contrarily, we detect homology in both species for the same family of blue viruses, leading us to create two independent endogenization events in SP6 and SP10 at two loci each (4 and 5). That the virus diversity has not yet been sampled or that the virus lineages disappeared could be two explanations for the monophyly of these two loci.



In this instance, which includes 3 clusters (and 3 distinct genes), 2 species are involved (SP4 and 5). According to species phylogeny, the two species SP4 and SP5 form a monophyletic group in the gene Cluster10. These two loci were likely endogenized independently of the other cases in the common ancestor of SP4 and SP5, and they were then attributed to the same event (6). In contrast, for Clusters 11 and 12, the two loci in SP4 and SP5 were probably edogenized in their common ancestor since they both share homology with the same green viral family, we then assign these EVEs to the same endogenization event (7).



In this example, there are only 1 species (SP3) and 2 clusters (so 2 different genes) involved. Since the loci SP3\_scf1,2,3, and 4 are likely multiple paralogs, we take a conservative approach to the number of EVEs and endogenization events and assume that they all result from the same endogenization event. We also placed another SP3 locus (\_scf5) within the same endogenization event because it exhibits homology with viruses in the yellow family (8).



In this example, which consists of three clusters (i.e., three distinct genes), there are EVEs in two species (SP1 and SP2). Due to their homologies to viruses in the blue family, the SP1 genome's EVE loci SP1\_scf1 and SP1\_scf2 are combined into a single endogenization event (9). The SP2\_scf1 and SP2\_scf2 loci, on the other hand, exhibit homologies with two distinct viral families (blue and green), respectively, and are not present in the same scaffold, so we count them as two separate endogenization events (10 and 11). We count SP1\_scf3 as a single event because it is the only locus in SP1 that shares homology with a family of green viruses (12).