Supplementary Information. Online Resource 2

Phylogenomic reconstruction indicates human-to-animal movement in pandemic *Klebsiella pneumoniae* lineages, with evidence of return in ST11

Journal: *Archives of Microbiology*

Authors: Dr Stephen Mark Edward Fordham[1], Dr Daniel Franklin[1] and Dr Elizabeth Sheridan[2]

Corresponding author: Stephen Mark Edward Fordham, Department of Life & Environmental Sciences, Bournemouth University, Poole, UK; sfordham@bournemouth.ac.uk

Affiliations:

[1]Bournemouth University, Department of Life & Environmental Sciences, Talbot Campus Fern Barrow, Poole BH12 5BB.

[2]University Hospitals Dorset NHS Foundation Trust, Department of Medical Microbiology, Poole Hospital, Longfleet road, BH15 2JB.

Contents

Supplementary Methods

Assessment of temporal signal

Temporal signal was assessed in two stages, and node ages were estimated only for clones that passed both. Stage 1 established whether the root-to-tip relationship contained more temporal information than expected under random assignments of the observed collection dates. Stage 2 tested whether population structure inflated the whole-tree substitution rate. The same analytical procedure was used throughout, allowing direct comparison of whole-tree and within-cluster estimates.

Stage 1. Whole-tree regression. For each clone, genetic divergence was regressed against sampling date on the clock-rerooted maximum-likelihood tree using TreeTime v0.12.1 (Sagulenko et al. 2018; Rambaut et al. 2016). Every isolate with a recorded collection date was included. The tree was rerooted by least squares with correction for shared ancestry. We report the coefficient of determination (R^2), the chi-squared statistic for the fit and the substitution-rate estimate with its standard error.

Reproducibility across random-number seeds. TreeTime resolves zero-length branches in the input tree stochastically. Each clone was therefore reanalysed using ten fixed seeds, comprising the integers 1 through 10. Consistency required recovery of the same root in all ten runs and variation in the substitution-rate estimate of less than 2% across seeds. Every clone met both requirements. The same root was recovered in 10 of 10 runs, and the maximum rate variation was 0.5% for ST307, 1.6% for ST11, 1.4% for ST15 and 1.3% for ST147.

Date-randomisation test. Reproducibility across seeds does not establish temporal signal. We therefore permuted collection dates among tips 100 times for each clone and rerooted the tree by least squares before re-estimating the root-to-tip regression for every permutation (Duchêne et al. 2015). These analyses generated a null distribution of R^2 values for date assignments containing no expected temporal structure. A clone passed stage 1 only when its observed R^2 exceeded the maximum R^2 among all valid randomised replicates. Permutations yielding no positively sloped root were recorded separately as having no valid clock. Rerooting each randomised dataset makes the test conservative by allowing every permuted date assignment to identify its best-fitting root (Table S2c).

Stage 2. Consistency of whole-tree and within-cluster rates. A whole-tree regression can be inflated by population structure because deep divergences between sublineages with different mean sampling dates may generate a date-divergence correlation without a molecular clock. Such inflation produces a whole-tree rate that exceeds rates estimated within individual SNP clusters. Informative clusters contained at least five dated genomes spanning at least three years. Their rates were combined into a precision-weighted estimate and compared with the whole-tree rate using a one-sided z-test at the 95% level. A clone failed stage 2 when its whole-tree rate significantly exceeded the pooled within-cluster estimate.

R^2 was not used as the stage-2 statistic because it depends on sampling window and sample size and does not directly measure clock consistency. Precision-weighted pooling also avoids multiple testing and prevents an imprecise rate from a small cluster from disproportionately influencing the conclusion. Cluster-specific estimates are reported in Table S2b.

ST307, ST11 and ST15 passed both stages and were assigned node-age estimates. Their observed R^2 values on the maximum-likelihood tree were 0.81, 0.62 and 0.49, respectively, whereas the randomised null distributions did not exceed 0.16, 0.11 and 0.10. All three also passed the within-cluster rate comparison. ST147 failed stage 1 because its observed R^2 of 0.25 fell within a randomised null distribution extending to 0.27. Its apparent signal also differed between the maximum-likelihood tree and the time-calibration fit, which produced R^2 values of 0.25 and 0.03, respectively. No node ages were therefore reported for ST147, which was presented on a midpoint-rooted genetic-divergence scale. Its within-cluster rate remained consistent with the whole-tree estimate, so it was rejected for lack of demonstrable temporal signal rather than for structure-related rate inflation.

Table S2a Temporal-signal assessment by clone. Panel A reports the final two-stage decision. Genomes are the total analysed and the number with a collection date. Whole-tree R^2 and rates are from the clock-rerooted time-calibration run. Stage 1 is determined by the date-randomisation test in Table S2c. The pooled cluster rate is the precision-weighted estimate from Table S2b. Positive z-values indicate a higher whole-tree rate, and 1.645 is the one-sided 95% threshold. Panel B reports reproducibility across seeds 1 through 10

Panel A. Temporal-signal and rate-consistency assessment

Clone	Genomes	Window	R^2 tree	Rate tree	Pooled cluster	z	Stage 1 DRT	Stage 2	Outcome
ST307	63 / 48	2014–2025	0.74	5.4×10^{-7}	7.4×10^{-7}	−1.01	pass	pass	datable
ST11	93 / 92	2008–2026	0.60	2.5×10^{-7}	1.9×10^{-7}	+0.78	pass	pass	datable
ST15	71 / 65	2008–2026	0.48	3.2×10^{-7}	2.9×10^{-7}	+0.40	pass	pass	datable
ST147	45 / 45	2020–2026	0.03	4.7×10^{-7}	5.3×10^{-7}	−0.18	fail	not tested	rejected (no temporal signal)

Panel B. Reproducibility across fixed random-number seeds

Clone	Seeds tested	Same root	Maximum rate variation
ST307	1–10	10 / 10	0.5%
ST11	1–10	10 / 10	1.6%
ST15	1–10	10 / 10	1.4%
ST147	1–10	10 / 10	1.3%

Table S2b Within-cluster root-to-tip regressions by SNP cluster. Every cluster containing at least five dated genomes spanning at least three years is shown. Rates are substitutions per site per year with one standard error. Cluster-specific z-values are shown for transparency, while the stage-2 decision used the pooled estimate in Table S2a

Clone	SNP cluster	n	Window	R^2	Rate	z
ST307	PDS000045377.17	9	2013.5–2019.5	0.60	$8.96 \pm 2.23 \times 10^{-7}$	−1.43

Clone	SNP cluster	n	Window	R^2	Rate	z
ST307	PDS000052254.9	6	2016.5–2023.1	0.93	$5.64 \pm 2.31 \times 10^{-7}$	−0.08
ST11	PDS000097502.32	43	2019.9–2026.5	0.31	$2.42 \pm 1.10 \times 10^{-7}$	+0.08
ST11	PDS000260965.1	33	2017.2–2024.5	0.44	$5.01 \pm 1.23 \times 10^{-7}$	−1.85
ST11	PDS000070340.11	14	2008.2–2023.5	0.15	$0.63 \pm 0.72 \times 10^{-7}$	+2.08
ST15	PDS000243746.2	40	2008.5–2021.5	0.31	$2.41 \pm 0.57 \times 10^{-7}$	+0.99
ST15	PDS000094658.3	5	2010.5–2016.5	0.92	$8.89 \pm 2.05 \times 10^{-7}$	−2.70
ST147	PDS000180176.26	43	2022.5–2026.5	0.19	$5.31 \pm 2.34 \times 10^{-7}$	−0.18

The smallest informative ST11 cluster, PDS000070340.11, was the only cluster that would exceed the stage-2 threshold if evaluated individually. Its rate estimate of $0.63 \pm 0.72 \times 10^{-7}$ substitutions per site per year had a standard error greater than the point estimate and was not distinguishable from zero. In the precision-weighted analysis, this imprecise estimate was outweighed by those from the larger clusters of 43 and 33 genomes, whose rates were consistent with the whole-tree estimate.

Table S2c Date-randomisation test of temporal signal. Collection dates were permuted among tips 100 times per clone. For every permutation, the maximum-likelihood tree was rerooted by least squares and the root-to-tip regression was re-estimated under the covariation-aware model. Permutations without a positively sloped root were recorded as having no valid clock. A clone passed when its observed R^2 exceeded the maximum value in the valid randomised null distribution

Clone	Observed R^2	Null R^2 median	Null R^2 95th	Null R^2 maximum	Permutations valid / no clock	Verdict
ST307	0.81	0.01	0.09	0.16	78 / 22	pass observed outside null
ST11	0.62	0.00	0.06	0.11	83 / 17	pass observed outside null
ST15	0.49	0.01	0.07	0.10	77 / 23	pass observed outside null
ST147	0.25	0.01	0.08	0.27	88 / 12	fail observed within null

Down-sampling sensitivity analysis

Discrete ancestral-state reconstruction uses the observed sample composition directly. Over-representation of a compartment within a cluster could influence the inferred transmission direction. This imbalance was pronounced in the North American companion-animal SNP cluster within ST11 (PDS000097502.32), which contained 33 of the clone's 38 animal genomes and 8 human genomes. To assess whether the inferred direction was an artefact of this imbalance, the 33 companion-animal genomes were randomly subsampled without replacement to retain 24, 16 or 8 genomes; the final level provided parity with the eight human genomes in the cluster. All other tips were retained. For each draw, the time-calibrated tree was pruned to the retained tips and the TreeTime mugration reconstruction was repeated. We performed 300 independent draws at each thinning level, giving 900 subsampled reconstructions in total (Figure S1).

For each replicate, we recorded (i) the reconstructed root (ancestral) compartment and (ii) whether the number of animal-to-human transitions was equal to or greater than the number of human-to-animal transitions on the pruned tree. At each thinning level, the reported proportion is the fraction of the 300 replicates that met the corresponding criterion.

Reproducibility was controlled at two levels so that the sampled subset was the only source of variation among replicates. A master random-number seed (42) fixed the sequence of animal subsets drawn across all replicates. Each mugration reconstruction was then assigned a fixed TreeTime seed, defined deterministically from the master seed and replicate index. In the absence of an explicit seed, TreeTime initialises its random-number generator from system entropy, rendering the reconstruction non-reproducible. Fixing the seed removes this source of stochastic variation; differences in transition counts therefore reflect the identities of the retained animal genomes rather than stochastic variation in TreeTime. We used 300 replicates per thinning level because a pilot analysis with 100 replicates per level yielded balanced-level proportions that varied by several percentage points among unseeded runs.

Treatment of tips of unknown compartment

Some isolates could not be assigned to either the human or animal compartment. ST11 included six such isolates: one environmental isolate and five isolates for which the host was not recorded, including two reservoir-clade genomes recovered from faeces without a documented host. ST307 included one isolate of unknown compartment.

These isolates were coded as unknown for the mugration reconstruction. TreeTime fits the discrete-trait model using the states observed in the data; the fitted model for both clones therefore comprised two states, animal and human. Unknown tips were treated as missing data rather than as a third state, and their compartments were inferred from the fitted model and their positions in the tree. Consequently, every tip in the reported reconstructions was assigned an animal or human state, and transitions involving tips of unknown observed compartment were counted in the same way as all other transitions. Thus, some reported transition counts incorporate model-inferred rather than observed compartment states.

To assess the influence of these inferred assignments, the reconstruction for each clone was repeated after pruning all unknown-compartment tips, leaving only isolates with recorded compartments. The pruned datasets comprised 87 of 93 ST11 genomes and 62 of 63 ST307 genomes. For both clones, pruning yielded the same inferred ancestral root state as the full reconstruction (human) and identical transition counts: four human-to-animal and six animal-to-human transitions for ST11, and five human-to-animal and two animal-to-human transitions for ST307. The inferred transmission direction for each clone was therefore unaffected by the tips for which compartment was inferred.

Because structured-coalescent analysis requires every sampled lineage to be assigned to a modelled deme, isolates with unknown compartments were excluded. This analysis included only isolates with both a known compartment and collection date: 87 of 93 genomes for ST11 and 47 of 63 genomes for ST307.

Structured-coalescent analysis

To determine whether the directional pattern persisted under a model that jointly estimates compartment-specific effective population sizes and migration rates, we fitted a marginal structured-coalescent model using MASCOT v3.0.7 (Müller et al. 2018) in BEAST v2.7.7 (Bouckaert et al. 2019). This was treated as a complementary sensitivity analysis, rather than as a direct correction for unequal sampling, because MASCOT jointly estimates effective population sizes and migration rates but does not model the sampling process. The analysis was applied to ST11, the clone most affected by sampling imbalance. The same model was fitted to ST307 as a specificity control because its discrete reconstruction showed the opposite directional pattern (an excess of human-to-animal transitions). A model that recovered direction from the phylogenetic data should therefore yield different directional estimates for the two clones.

Model specification and priors. The model included isolates with both a known compartment and a collection date (87 ST11 isolates and 47 ST307 isolates). Animal isolates were not subdivided into livestock and companion-animal compartments because neither subgroup was large enough to support a separate effective-population-size estimate. Analysis of variable sites alone would incorrectly imply that the entire genome evolves at the rate observed at polymorphic sites and thereby inflate the substitution-rate estimate. Counts of invariant A, C, G and T positions in the recombination-masked core alignment were therefore supplied as constant-site weights alongside the variable sites (Leaché et al. 2015). For ST11, the invariant-site counts were A, 840,855; C, 1,162,527; G, 1,163,142; and T, 841,278 (634 variable sites in a 4.01-Mb core). Corresponding counts for ST307 were A, 997,274; C, 1,359,615; G, 1,359,700; and T, 995,061 (572 variable sites in a 4.71-Mb core). Substitutions were modelled using GTR with estimated base frequencies, a strict molecular clock and a constant effective population size within each compartment.

Priors. Effective population sizes were assigned independent lognormal priors with a mean of 2.0 and standard deviation of 2.0 on the natural-log scale (median, 7.4; approximate 95% prior interval, 0.15–370). Migration rates were assigned identical exponential priors with a mean of 1.0 in both directions, providing a directionally symmetric migration prior. The clock rate was assigned a lognormal prior with a mean of −13.8155 and standard deviation of 1.5 on the natural-log scale. This corresponds to a median of 1.0×10^{-6} substitutions per site per year, the published *Klebsiella pneumoniae* core-genome substitution rate (Bowers et al. 2015) and an approximate 95% prior interval of 5×10^{-8} to 2×10^{-5}. GTR relative rates were assigned standard diffuse Gamma priors (shape, 0.05), and base frequencies were assigned a uniform Dirichlet prior.

Direction convention and migration flux. MASCOT parameterises its internal migration matrix backwards in time. To derive forward-time transmission rates, the model was specified using a forwards-migration parameter. MASCOT converts this parameter to the backward-time matrix internally (m_backward[a][b] = f[a→b] × Ne[b] / Ne[a]) and logs rates with the deme indices reversed. Therefore, a logged rate labelled Human_to_Animal represents the forward-time human-to-animal transmission rate

and can be compared directly with the mugration transition counts. Directionality was summarised using both migration rate and migration flux. The migration rate is the forward-time rate of movement per lineage between compartments. Migration flux is the rate multiplied by the effective population size of the source compartment (rate × Ne of the source deme), thereby accounting for the dependence of the per-lineage rate on source-population size and estimating the number of lineages moving per unit time. For each measure, we report the posterior median and 95% highest posterior density (HPD) interval in both directions, together with the posterior probability that the animal-to-human value exceeds the human-to-animal value.

Chains and convergence. Chains were sampled every 5,000 states and continued until all model parameters achieved an effective sample size (ESS) >200 in Tracer v1.7.2 (Rambaut et al. 2018). The first 10% of each chain was discarded as burn-in. The ST11 chain comprised 24.8 million states, with a minimum post-burn-in ESS of 821 (clock rate); the ST307 chain comprised 45.5 million states, with a minimum ESS of 4,671.

Supplementary Results

Structured-coalescent migration estimates

Posterior estimates for both clones are presented in Table S3. The fitted model favoured animal-to-human movement for both clones. For ST11, the posterior probability that the animal-to-human rate exceeded the human-to-animal rate was >0.999; the corresponding probability for migration flux was 0.995. For ST307, these probabilities were 0.982 and 0.921, respectively.

The specificity control indicates that this shared directional preference should not be interpreted as independent confirmation of transmission direction. The discrete mugration reconstruction distinguished the clones: ST307 showed an excess of human-to-animal transitions (five versus two), whereas ST11 showed an excess of animal-to-human transitions (six versus four). In contrast, the structured-coalescent model favoured animal-to-human movement for both clones. This pattern is consistent with the estimated effective-population-size asymmetry: the median human effective population size was substantially greater than the animal estimate in both ST11 (47.0 versus 3.8) and ST307 (19.8 versus 3.0). Given the small estimated animal deme, the model coalesced animal lineages rapidly and favoured forward-time movement from animals to humans, largely irrespective of the locations of sampled genomes in the tree. Under this specification, the structured-coalescent result was therefore dominated by effective-population-size asymmetry and did not independently resolve transmission direction. We consequently interpret it as a sensitivity analysis whose specificity control did not meet its intended criterion. Directional inference in the main text is based on the discrete reconstruction, tree rooting and collection dates (Table 2), rather than on MASCOT.

This directional preference was not attributable to the migration prior, which was identical in both directions (exponential, mean 1.0). Posterior migration rates differed by approximately two orders of magnitude in each clone (Table S3), indicating that the separation was informed by the data rather than imposed by the prior. However, the estimates were sensitive to the effective-population-size posteriors, and formal sensitivity analyses using alternative effective-population-size priors were not performed. We therefore present the structured-coalescent analysis as a sensitivity check rather than as confirmatory evidence of transmission direction.

Table S3 Structured-coalescent (MASCOT) posterior estimates for ST11 and the ST307 specificity control. Migration rates are forward-time rates per lineage per year. Migration flux is the rate multiplied by the effective population size of the source deme. Values are posterior medians with 95% highest posterior density (HPD) intervals. Ne denotes effective population size. P(A→H > H→A) denotes the posterior probability that the animal-to-human value exceeds the human-to-animal value

Parameter (forward-in-time)	ST11	ST307 (control)
Isolates (human / animal)	87 (49 / 38)	47 (29 / 18)
Ne, human	47.0 (25.8–75.5)	19.8 (5.8–36.6)
Ne, animal	3.8 (1.5–7.1)	3.0 (0.7–9.6)
Migration rate, human → animal	0.013 (0.002–0.030)	0.029 (0.001–0.118)
Migration rate, animal → human	1.29 (0.12–3.34)	0.68 (0.03–2.28)
P(rate A→H > H→A)	> 0.999	0.982
Migration flux, human → animal	0.61 (0.12–1.25)	0.57 (0.04–1.74)
Migration flux, animal → human	4.83 (0.65–10.35)	2.11 (0.12–6.37)
P(flux A→H > H→A)	0.995	0.921
Clock rate (subs/site/yr)	3.8×10^{-7}	6.2×10^{-7}
tMRCA, median (95% HPD)	1976 (1965–1986)	2002 (1996–2006)
Chain length / minimum ESS	24.8M / 821	45.5M / 4,671

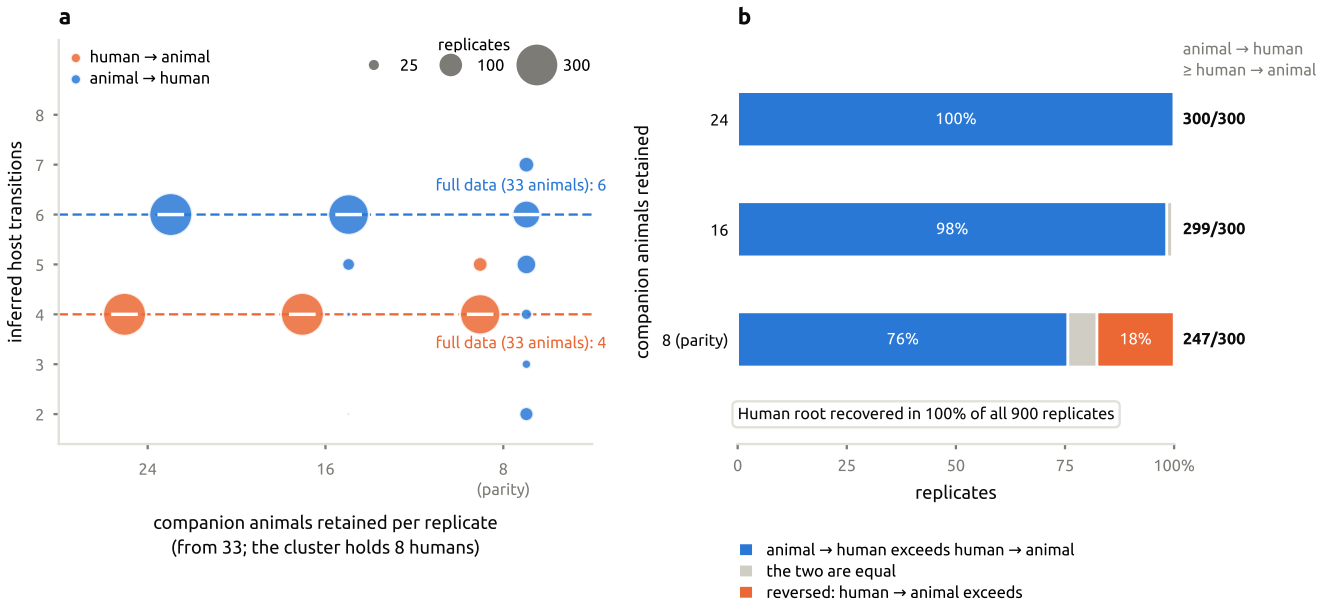

Fig. S1 The direction of transmission inferred for ST11 is not an artefact of companion-animal oversampling. Thirty-three of ST11's 38 animal genomes lie in a single companion-animal SNP cluster (PDS000097502.32), alongside 8 human genomes. Animal genomes in this cluster were randomly subsampled to 24, 16 or 8 genomes, with the final level matching the eight human genomes. All other tips were retained, and the mugration reconstruction was repeated on each pruned tree for 300 independent subsamples at each level (900 in total)

(a) Host transitions inferred in each replicate, coloured by destination compartment (orange, human-to-animal; blue, animal-to-human) as in the main figures. Circle area is the number of replicates returning that count; white bars mark medians. Dashed lines give the values obtained from the full data (4 human-to-animal, 6 animal-to-human). The medians are unchanged at every thinning level, and the distributions widen rather than shift as animals are removed.

(b) The proportion of replicates in which animal-to-human transitions exceeded, equalled or fell below human-to-animal. The bold figures give the proportion in which the animal-to-human excess was retained or tied. The excess survives balanced sampling of the cluster in the great majority of replicates; the human root is recovered in every replicate at every level.