

HRSV classification below the subgroup level

1 **Title: The unified proposal for classification of human respiratory syncytial virus below the**
2 **subgroup level.**

3 Stephanie Goya 1^{#1}, Christopher Ruis 2, 3, 4², Richard A. Neher 5², Adam Meijer 6, Ammar Aziz 7,
4 Angie S. Hinrichs 8, Anne von Gottberg 9,10, Cornelius Roemer 5, Daniel G. Amoako 11³, Dolores
5 Acuña 12, Jakob McBroome 8,13, James R. Otieno 14, Jinal N. Bhiman 9,10, Josie Everatt 9, Juan C.
6 Muñoz-Escalante 15, Kaat Ramaekers 16⁴, Kate Duggan 17, Lance D. Presser 6, Laura Urbanska 5,
7 Marietjie Venter 18, Nicole Wolter 9,10, Teresa C. T. Peret 19, Vahid Salimi 20, Varsha Potdar 21,
8 Vítor Borges 22, Mariana Viegas 12,23^{#1}

9
10 1. Department of Laboratory Medicine and Pathology, University of Washington Medical Center,
11 Seattle, WA, USA

12 2. VPD Heart Lung Research Institute, University of Cambridge, Cambridge, UK

13 3. Cambridge Centre for AI in Medicine, University of Cambridge, UK

14 4. Department of Veterinary Medicine, University of Cambridge; Cambridge, UK

15 5. Biozentrum, University of Basel and SIB, Basel, Switzerland

16 6. Centre for Infectious Disease Control, National Institute for Public Health and the Environment,
17 Bilthoven, The Netherlands

18 7. WHO Collaborating Centre for Reference and Research on Influenza, Victorian Infectious Diseases
19 Reference Laboratory, The Royal Melbourne Hospital, Parkville, Australia

20 8. Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA

21 9. National Institute for Communicable Diseases of the National Health Laboratory Service,
22 Johannesburg, South Africa

23 10. University of Witwatersrand, Johannesburg, South Africa

24 11. Antimicrobial Research Unit, College of Health Sciences, University of KwaZulu-Natal, Durban,
25 South Africa

26 12. Laboratorio de Salud Pública, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La
27 Plata, Buenos Aires, Argentina.

28 13. Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA,
29 USA

30 14. Theiagen Genomics, Highlands Ranch, CO, USA.

31 15. Center for Research in Health Sciences and Biomedicine, Autonomous University of San Luis
32 Potosí, Mexico

¹ These authors were co-principal investigators.

² These authors contributed equally to this article.

³ current address: Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

⁴ current address: Sciensano, Infectious diseases in humans, unit (re)-emerging viruses, Brussels, Belgium

HRSV classification below the subgroup level

33 16. KU Leuven, Department of Microbiology, Immunology and Transplantation, Rega Institute for
34 Medical Research, Laboratory of Clinical and Epidemiological Virology, Leuven, Belgium.

35 17. Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, United Kingdom

36 18. Center for Emerging Arbo and Respiratory virus research, Department Medical Virology,
37 University of Pretoria, Pretoria, South Africa

38 19. Galveston National Laboratory, Department of Microbiology and Immunology, University of
39 Texas Medical Branch, Galveston, Texas, USA.

40 20. Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran,
41 Iran

42 21. ICMR National Institute of Virology, Pune, India

43 22. Genomics and Bioinformatics Unit, Department of Infectious Diseases, National Institute of
44 Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

45 23. National Scientific and Technical Research Council, Buenos Aires, Argentina.

46

47 **Abstract**

48 A globally implemented unified classification for human respiratory syncytial virus (HRSV) below
49 the subgroup level remains elusive. Here, we formulate the global consensus of HRSV classification
50 based on the challenges and limitations of our previous proposals and the future of genomic
51 surveillance. From a high-quality dataset of 1,480 HRSV-A and 1,385 HRSV-B genomes submitted
52 to NCBI and GISAID up to March 2023, we categorized HRSV-A/B sequences into lineages based on
53 phylogenetic clades and amino acid markers. We defined 24 lineages within HRSV-A and 16 within
54 HRSV-B, providing guidelines for prospective lineages definition. Our classification demonstrated
55 robustness in its applicability to both complete and partial genomes. In addition, it allowed the
56 observation of notable lineage replacements and the identification of lineages exclusively detected
57 since the COVID-19 pandemic. We envision that this unified HRSV classification proposal will
58 strengthen and facilitate HRSV molecular epidemiology on a global scale.

59

HRSV classification below the subgroup level

60 **Introduction**

61 Significant progress has been achieved in developing interventions against the human respiratory
62 syncytial virus (HRSV), a leading cause of acute lower respiratory tract infection in children, elderly,
63 and immunocompromised individuals. In 2023, the U.S. Food and Drug Administration (FDA) and
64 the European Medicines Agency (EMA) approved the first HRSV vaccines (1,2). Simultaneously, a
65 monoclonal antibody was approved for widespread use in infants and not limited to high-risk and
66 premature children (3). The availability of HRSV immunization highlights the role of molecular
67 epidemiology as a tool to monitor their efficacy. Nevertheless, a standardized HRSV genotype system
68 has yet to be defined and implemented.

69

70 In 2022, HRSV was designated as "*Orthopneumovirus hominis*" species within the *Pneumoviridae*
71 family. Below species level, there are two antigenic groups known as HRSV subgroup A (HRSV-A)
72 and B (HRSV-B), that were previously referred to as subtypes (4–6). Within each subgroup,
73 genotypes were initially defined based on statistically supported phylogenetic clades inferred with the
74 second hypervariable region (2HR) of the G gene (Figure 1A-B) (7). The G gene, encoding the
75 attachment glycoprotein, exhibits the highest genetic and antigenic variability. Notably, this gene has
76 undergone a duplication of a 72-nt fragment in HRSV-A and 60-nt in HRSV-B (Figure 1B) (8,9).

77 To identify emerging genotypes, genetic distances between phylogenetic clades and distinctive
78 genetic features were used, accompanied by variable nomenclature based on the gene (GA1–GA7 in
79 HRSV-A and GB1–GB4 in HRSV-B), country and subgroup (SAB1–SAB4 for South African
80 genotypes in HRSV-B) or city/province (NA1-NA2 [Niigata], ON1 [Ontario] in HRSV-A and BA1-
81 BA9 [Buenos Aires] in HRSV-B) (7–15). Since 2019 alternative phylogenetic reclassifications were
82 proposed using G gene sequences or complete genomes. From one side, the use of the G gene allows
83 laboratories without capacity for whole genome sequencing (WGS) to perform molecular
84 epidemiology studies (16,17). From the other side, complete genomes allow the surveillance of all
85 viral genes, including the F gene encoding the fusion protein, which is the target of monoclonal
86 antibodies, and the basis of approved vaccines against HRSV (Figure 1A) (18,19).

87 The milestones in HRSV interventions have renewed interests to address the challenge of classifying
88 HRSV below the subgroup level. This prompted establishment of the HRSV Genotyping Consensus
89 Consortium (RGCC), formed by HRSV and virus evolution experts aiming to standardize the HRSV
90 classification system. Here, we present a framework for HRSV classification below the subgroup
91 level, based on current knowledge of HRSV diversity and evolution, focused on practical
92 implementation for molecular epidemiology.

93

94 **Methodology**

95 **HRSV sequences dataset.**

HRSV classification below the subgroup level

96 The HRSV complete genomes available from NCBI GenBank
97 (<https://www.ncbi.nlm.nih.gov/labs/virus>) and GISAID EpiRSV (<https://gisaid.org/>) up to March 11,
98 2023 were downloaded using a filter for sequence length above 14,000-nt, obtained from human hosts
99 and including the year and country of the sample collection. Sequences were categorized into files
100 based on their reported subgroup resulting in 2,744 HRSV-A and 2,443 HRSV-B genomes
101 (Supplementary Figure 1). Sequences containing nucleotide ambiguities, indicative of inadequate
102 sequencing depth, were reserved for epidemiological analysis but excluded from formal lineage
103 definition. Removal of genomes with any nucleotide ambiguities was performed using BBmap
104 (reformat, version Jan-2021). To ensure diversity without redundancy, BBmap (dedupe, version Feb-
105 2020) was used to remove identical sequences, preserving one representative.
106 Sequences were aligned with MAFFT v7.490, and alignments were inspected and corrected with
107 Aliview v1.28, mainly in the G gene (20,21). Furthermore, alignment ends were trimmed to
108 encompass complete genomes from the first codon of the first gene (NS1) to the last codon of the last
109 gene (L). Partial genomes were considered if the lack of sequence was within 50-nt of the genome
110 ends. Genomes with nucleotide insertions or deletions causing frameshift in any open reading
111 frame(s) were identified with RSVsurver and removed (22). Following alignment trimming, the
112 presence of identical sequences with nucleotide differences in the trimmed region prompted another
113 round of redundancy removal using BBmap tool, resulting in the final set of 1,538 HRSV-A and
114 1,387 HRSV-B genomes (Supplementary Figure 1).

115

116 **Phylogenetic analysis.**

117 Maximum likelihood phylogenetic trees were constructed with IQ-TREE v2.2.0 (details in
118 Supplementary Material). Monophyletic clades were considered statistically supported when SH-
119 aLRT value was $\geq 80\%$ and UFBoot2 value was $\geq 90\%$ (23,24). Phylogenetic trees were visualized
120 with Figtree v1.4.4 and Auspice. Temporal signal was assessed with TempEst v1.5.3, and molecular-
121 clock phylogenies were inferred with TreeTime (25).

122 The ancestral sequence reconstruction was inferred using Augur bioinformatic toolkit v23.1.0 (26).
123 Recombination events were assessed with alignment-based method using RDP4 and phylogenetic-
124 based TreeKnit (Supplementary Material).

125 The amino acid substitutions linked to the clades in the tree were inferred using Augur and the initial
126 screening of lineages was automated with Autolin (27). Amino acid comparison among monophyletic
127 clusters underwent manual curation to rectify conflicts arising from internal (nested) lineages and the
128 confirmation of the lineage-defining amino acids in more than 90% of the clade's sequences. Results
129 from this study are available: https://github.com/rsv-lineages/Classification_proposal.

130

131 **Results and Discussion.**

132 **Baseline agreements on the HRSV classification definition.**

HRSV classification below the subgroup level

133 The proposed classification establishes “HRSV lineages” for viruses below subgroup level. Recent
134 studies have demonstrated superior resolution provided by complete genomes in constructing HRSV
135 phylogenetic trees (16,18,19). Anticipating improvement of WGS, we defined a classification system
136 based on maximum likelihood phylogenetic trees inferred from complete HRSV genomes. The
137 maximum likelihood algorithm formulates hypotheses about the evolutionary relationships among
138 sequences, and the implementation within IQ-TREE dealing with large datasets makes it particularly
139 well-suited to assert HRSV genomic phylogeny including sequences collected >50 years ago (23).

140

141 The definition of complete genome used in this study spans from the first codon of the first gene
142 (NS1) to the last codon of the last gene (L). “Almost” complete genomes were considered if the
143 sequence information gaps were within a 50-nt window at the genome ends. For a robust genetic
144 dataset, only genomes without nucleotide ambiguities (according to the IUPAC code for nucleotide
145 degeneracy) were used to define lineages.

146

147 **Composition of the genomic dataset used for lineage classification.**

148 Applying the established baseline agreements, we gathered 1,538 HRSV-A and 1,387 HRSV-B high-
149 quality, public genomes. Analysis of the genomic dataset revealed a limited global HRSV genomic
150 surveillance until 2007, with <20 genomes deposited annually (Figure 2A, Supplementary Figure 2).
151 Since 2008, there was improvement in the number of genomes and representation of countries,
152 probably attributed to increased adoption of sequencing technology. Additionally, a significant surge
153 occurred post-2021, probably driven by expansion of viral genomics since the SARS-CoV-2
154 pandemic and the approval of the HRSV prophylactic treatments (Figure 2A, Supplementary Figure
155 2). Considering delays in genome deposition in public databases, the number of genomes in 2022 may
156 rise if the submission dates to the genomic databases postdate the data used in this study. Regarding
157 geographic representation, nine countries (Australia, United Kingdom, New Zealand, USA,
158 Argentina, Kenya, Morocco, Netherlands, and Brazil) submitted >100 genomes, only United
159 Kingdom achieved uninterrupted surveillance since 2008, but Australia deposited the most genomes
160 globally (Figure 2B).

161

162 **Defining the root and the final dataset of HRSV phylogenetic trees.**

163 We reconstructed separate maximum likelihood phylogenetic trees for the HRSV-A/B datasets. The
164 first step toward a reliable classification system requires precise determination of the phylogenetic
165 tree root. Two approaches were used: 1) utilization of an outgroup, a conventional method for
166 inferring the tree root using sequences known to be evolutionarily distant, and 2) phylodynamic
167 analysis, integrating temporal and phylogenetic patterns in virus evolution (Supplementary Material,
168 Supplementary Figure 3). Both approaches consistently identified the same root for each subgroup
169 cluster. Phylodynamic analysis also identified 58 outlier sequences for HRSV-A and two for HRSV-B

HRSV classification below the subgroup level

170 that were excluded from lineage designation. The final dataset considered for lineage designation
171 comprised 1,480 HRSV-A and 1,385 HRSV-B genomes (Supplementary Table 1 and 2).

172

173 **The HRSV lineage definition.**

174 An HRSV lineage was defined as a statistically supported monophyletic cluster comprising ≥ 10
175 sequences and characterized by ≥ 5 amino acid substitutions compared to the parental lineage. The
176 lineage-defining amino acids, present in $\geq 90\%$ of the sequences within the clade, may be found in any
177 of the viral proteins.

178 Phylogenetic classifications vary among viral species, and all try to ensure the defined clustering
179 reflects the heterogeneity of the viral population, therefore requiring the consideration of unique
180 evolutionary characteristics of each viral species, the use of arbitrary thresholds and the feasibility for
181 long-term use (28–30). There is inherent bias in any classification due to sequence availability and
182 spatiotemporal representation of viral diversity. Therefore, HRSV lineage definition does not include
183 criteria of sequences from different outbreaks or countries to ensure early detection of new lineages.

184 We observed that presence of distinctive amino acids shared by sequences of a phylogenetic clade in
185 comparison to the parental lineage is a simple method to identify a new lineage. Methods (i.e. average
186 nucleotide genetic distances, average patristic distances, or patristic distances between nodes) need
187 phylogenies with complete datasets to define new categories, becoming complex with rapid increases
188 of available sequences (16–19). In this proposal, different amino acid thresholds were initially
189 screened in an automated manner, ranging from 1-10 lineage-defining amino acids (Supplementary
190 Material). The number of small lineages decreases as the number of lineage-defining amino acids
191 increases, and five amino acids resulted in an intermediate complexity of lineages defined for both
192 HRSV subgroups. We acknowledge that other thresholds could be useful, but we emphasize that the
193 key to establishing a global consensus is clear operational guidelines and a robust classification, two
194 aspects that our proposal fulfills.

195

196 **a) Definition of the lineage's nomenclature**

197 Lineage nomenclature integrates the HRSV subgroup letter and ascending ordinal numbers, separated
198 by dots to represent nested lineages (Figure 3A-B). Furthermore, a distinct nomenclature was
199 assigned to the 72-nt (24 amino acids) and 60-nt (20 amino acids) G-gene duplication within HRSV-
200 A/B, respectively. This genetic event is epidemiologically relevant, as viruses with this duplication
201 have exclusively been detected since 2016 (31–34). To facilitate tracking these viruses, we use the
202 alias "D", specifically A.D (historically, ON1 genotype) and B.D (historically, BA genotype) for
203 HRSV-A/B respectively, and nested lineages with increasing ordinal numbers.

204 In summary, letters A and B indicate the HRSV subgroup at the beginning of the lineage name, C is
205 unused, and D serves as an alias for 72-nt and 60-nt duplication within the G gene. In addition, aliases
206 starting from E are limited to three numerical levels of nested lineages, preventing indefinite

HRSV classification below the subgroup level

207 accumulation of numbers. For example, B.D.4.1.1 lineage has descendant lineages named B.D.E.1 -
208 B.D.E.4 instead of B.D.4.1.1.1 - B.D.4.1.1.4, where “E” represents “4.1.1” (Figure 3A-B). The
209 nomenclature is based on the tree topology, reflecting the order of the nodes from the root to the tips,
210 but it is unrelated to the sequence collection date or date of the most recent common ancestor of the
211 lineage.

212 To remain functional, a nomenclature system requires periodic updates as new lineages emerge.
213 Therefore, we have established two open repositories on GitHub containing definitions of each
214 lineage, signature mutations, and representative sequences. These repositories, available at
215 <https://github.com/HRSV-lineages/lineage-designation-A> and [https://github.com/HRSV-](https://github.com/HRSV-lineages/lineage-designation-B)
216 [lineages/lineage-designation-B](https://github.com/HRSV-lineages/lineage-designation-B) are intended to provide up-to-date definitions, and serve as a platform
217 for discussion and designation of novel lineages.

218

219 **b) Identification of lineages within the HRSV-A and HRSV-B rooted trees.**

220 We reconstruct ancestral sequences at the root of the phylogenetic trees. While these sequences are
221 not biologically real, they serve as surrogate parental lineages during initial classification. Identifying
222 monophyletic clusters with ≥ 10 sequences and ≥ 5 amino acid changes compared to the reconstructed
223 root sequence, we defined three HRSV-A lineages (A.1 – A.3) and four HRSV-B lineages (B.1 –
224 B.4). Two sequences, (EPI-ISL-15771600_USA_1956 and MG642074_USA_1980) could not be
225 classified perhaps because they belong to underrepresented extinct lineages.

226 The first lineages underwent further analysis to identify nested lineages in an iterative manner,
227 resulting in a total of 24 lineages within HRSV-A, and 16 within HRSV-B (Figure 3 and Figure 4).

228 Close to the root of the HRSV-B tree, extinct lineages were underrepresented, comprising < 10
229 sequences but featuring > 5 distinct amino acids (B.1, B.3, B.4). Despite the low number of sequences,
230 we included them as lineages to trace evolutionary branches that gave rise to currently circulating
231 lineages. In addition, A.D.2 is slightly below the sequences threshold, but we kept the lineage
232 category to emphasize the common ancestor among A.D.2.1 and A.D.2.2.

233 We scrutinized the presence and absence of the duplication in the G gene across each tree. While
234 patterns were mostly as expected with a single historical duplication event, some genomes within the
235 clade with the duplication in G, lacked the duplication. The dispersed association of these sequences
236 in the phylogenetic tree, rather than forming a monophyletic cluster, suggests the virus did not lose
237 the nucleotide duplication (Supplementary Figure 4). Instead, similar read length to the duplication
238 region of certain short-read NGS technologies potentially masks the presence of the duplication when
239 used in the consensus genome assembly with reference sequences that do not possess the nucleotide
240 duplication. Therefore, we recommend the use of NGS data with quality filtered reads with a length
241 greater than 150-nt to avoid this problem.

242 Lineage-defining amino acids were present in all HRSV proteins, the majority identified within the G
243 protein (Tables 1 and 2). Also, the polymerase L protein was noteworthy, contributing to the

HRSV classification below the subgroup level

244 distinction of 21 out of 24 HRSV-A lineages and 15 out of 16 HRSV-B lineages (Table 1 and Table
245 2). Interestingly, the F protein contributed to define 14 lineages in HRSV-A and 13 in HRSV-B
246 (Figure 3B, Figure 4B). Both the G and F surface glycoproteins are under selection pressure from
247 humoral immunity and exhibit a robust phylogenetic signal (18,32). While the G protein displays
248 substantial nucleotide and amino acid sequence plasticity, the F protein experiences strong negative
249 selection, likely attributed to functional/structural constraints (35). In fact, the fusion peptide is the
250 only region in F without lineage-defining amino acids (Figure 3B, Figure 4B). While the low diversity
251 of the F protein is promising for HRSV interventions, monitoring the F protein during their global
252 implementation is essential.

253

254 **The use of G sequences with the HRSV lineage classification system.**

255 The main challenge for global expansion of HRSV genomics is the absence of a cost-effective,
256 globally standardized and validated methodology for sequencing, in contrast to SARS-CoV-2 or
257 influenza virus (36,37). In addition, limited funding and/or infrastructure causes some laboratories to
258 prefer sequencing the G gene only (38–40). While we recommend the use of complete genomes for
259 HRSV lineage assignment, partial genomes covering the G and/or F genes can be used since they
260 reproduce the topology of the HRSV tree (16,18). As previous reports showed the decreased
261 phylogenetic signal, we do not recommend the use of smaller G gene regions such as the second
262 hypervariable region (250-nt length at the 3' gene end, Figure 1) that was used historically for
263 molecular epidemiology (16).

264 Minimal misclassification (1.2% error) was found in HRSV-A and none in HRSV-B when we
265 explored the reproducibility of the classification using the G gene only, retaining strong support for
266 lineage-defining nodes (Supplementary Material). However, the use of G gene limits the detection of
267 emergent lineages and precludes monitoring of potential immune evasion in the F protein.

268

269 **Prospective HRSV lineage assignment and definition.**

270 The assignment of sequences to the existing lineages can be automated using online tools such as
271 NextClade (<https://clades.nextstrain.org/>) (41), ReSVidex (<https://cacciabue.shinyapps.io/resvidex/>),
272 INSaFLU (<https://insaflu.insa.pt>) (42) or UShER (<https://usher.bio/>) (43). However, for a classical
273 approach to define the lineage of query sequences, we encourage users to follow our guidelines
274 described in the Supplementary Material and available on GitHub ([https://github.com/orgs/rsv-
275 lineages/repositories](https://github.com/orgs/rsv-lineages/repositories)).

276 We anticipate new lineages of HRSV-A/B will continue to emerge, and we envision expanding our
277 proposed nomenclature to incorporate new lineages. Detection and definition of a new lineage should
278 adhere to the considerations outlined in this study and described in our operative guidelines
279 (Supplementary Material).

HRSV classification below the subgroup level

280 It is important to stress that assigning the lineage of a sequence does not require the completeness of
281 the genome or the absence of nucleotide ambiguities, it only requires supported association in a
282 phylogenetic tree. In this sense, interpretation of results when complete genomes are not used should
283 consider the intrinsic limitations. However, to define a new lineage, complete genomes and the
284 absence of ambiguities are mandatory due to the amino acid characterization as part of the definition.
285 We encourage the general community to report detection of new HRSV lineages at the RGCC GitHub
286 page as an 'issue' within the corresponding repository for HRSV-A/B. The RGCC study group will
287 evaluate the newly proposed lineage, and if confirm, reference alignments will be updated.

288 **Molecular epidemiology of HRSV with the proposed classification.**

289 Using our proposal, we described the HRSV molecular epidemiology including all available genomes,
290 even those previously discarded due to reasons such as the presence of nucleotide ambiguities.
291 Consequently, we analyzed the seasonality of lineages using a dataset comprising 2,277 HRSV-A and
292 2,058 HRSV-B genomes, revealing notable co-circulation and lineage replacement over time (Figure
293 3C, Figure 4C). Within HRSV-A, A.1 and A.2 lineages are currently extinct, with the last detected
294 sequences collected in 1995, and 2015, respectively. Since 2011, A.D and nested lineages continue to
295 circulate. Descendant lineages within A.D (A.D.2.2 and A.D.4) were first detected in 2013,
296 highlighting the rapid divergence of the HRSV-A viruses with the duplication in G gene. Within
297 HRSV-B, the circulation of lineages B.1, B.2, B.3, and B.4 exhibited a strong lineage replacement
298 (Figure 4C). Although the first reported detection of viruses with a 60-nt duplication in the G gene
299 (B.D lineage) traces back to 1999, complete genomes only became available from 2005 onwards (8).
300 By 2009, only genomes from B.D and nested lineages were detected, and from 2017 onwards, only
301 B.D.4 and nested lineages have been observed.

302 Post-COVID-19 pandemic, a higher number of HRSV-A than HRSV-B lineages were detected,
303 especially from A.D.1, A.D.3-A.D.5 divergent branches, while A.D.2 exhibited sporadically low
304 detections. Post-pandemic HRSV-B viruses primarily classified within lineages B.D.4.1 and its
305 divergent branches, B.D.4.1.1 and descendants B.D.E.1 – B.D.E.5. While this is an overall analysis
306 about post-pandemic HRSV spread, further analysis is needed to understand global transmission, local
307 evolution, and detect geographically limited lineages due to SARS-CoV-2 control measures. Pre-
308 pandemic lineages may not have been described due to weak genomic surveillance. However, if pre-
309 or post-pandemic HRSV genomes are shared publicly in the future, our classification will allow
310 updates via RGCC GitHub when a new lineage is discovered.

311 With the current data, we observed that some lineages were only post-pandemic -A.D.3.1, A.D.5.2,
312 and A.D.5.3 in HRSV-A, and B.D.E.1 and B.D.E.3 in HRSV-B. However, most of the lineage-
313 defining amino acids in the new lineages were found pre-pandemic. For example, A.D.5.2 acquired

HRSV classification below the subgroup level

314 lineage status when the substitution C26Y in M2-2 appeared, while the other four substitutions were
315 already present in the closest genome of the parental lineage from 2019 (GenBank accession number
316 MZ515825). It is important to mention the detection of the new post-pandemic lineages does not
317 contradict recent studies stating the absence of new post-pandemic 'genotypes' as they were based on
318 previous classification systems (44–47). In addition, the possibility that these new post-pandemic
319 lineages have been circulating since before the pandemic depends on the deposition of genomes.

320

321 **Final remarks and conclusion.**

322 A consensus classification of HRSV below the subgroup level has been a challenge for multiple
323 decades. Collaboratively, the HRSV molecular evolution research community, along with experts in
324 the evolution of other respiratory viruses, have worked toward establishing a unified global
325 classification system. This proposal categorizes HRSV-A/B sequences into lineages based on
326 phylogenetic associations and amino acid markers, relying on complete genomes. While defining a
327 new lineage requires the complete genome sequence, partial or low-quality genomes can be assigned
328 to the existing lineages, emphasizing the robustness of this system. We developed standard guidelines
329 for the lineage definition, lineage assignment, and created online resources for updates, ensuring long-
330 term utility.

331 Defining a viral category below species through a phylogenetic-based classification is challenging.
332 The system must exhibit reproducibility, balance complexity, and be updatable to capture the level of
333 heterogeneity useful for viral surveillance. While different proposals may emerge, they may not
334 succeed in their ability to address the specific questions for a given virus. Illustratively, three SARS-
335 CoV-2 phylogenetic-classification were proposed by Nextstrain, GISAID, and Pango (48,49).
336 However, Pango classification proved to be instrumental to monitor a rapidly evolving virus within a
337 constrained temporal window and dealing with the meticulous detection of emergent mutations
338 through unparalleled global genomic surveillance efforts.

339 HRSV is not an emerging virus; it generates annual outbreaks with co-circulation and replacement in
340 the prevalence of its antigenic subgroups. Though few, there are RSV genomes collected from clinical
341 samples collected >50 years ago. The largest increase in the number of genomes occurred in the last
342 two years, and we anticipate this trend continuing. In 2023, recombinant F protein vaccines were
343 approved for the first time. As their implementation progresses, we will learn how this affects viral
344 evolution. We hope this unification proposal for the phylogenetic classification of HRSV will
345 facilitate spatiotemporal comparative lineage surveillance and detection of emerging lineages. A
346 limitation of our definition is the uncertainty regarding the impact of individual amino acid
347 substitutions on lineages. We do not know a priori whether a simple amino acid substitution may have
348 a significant epidemiological impact. Hence, whole genome surveillance and the study of the lineage-
349 phenotype association will be essential, as observed in genetic and antigenic characterization in
350 influenza to estimate the effectiveness of immunization (50). Other knowledge gaps include, but are

HRSV classification below the subgroup level

351 not limited to, analyzing the association between lineages and the severity of HRSV disease, as well
352 as associations of particular lineages with demographic characteristics of individuals.

353

354 **Acknowledgments**

355 Authorships between the second and last were listed alphabetically by name. We acknowledge the
356 authors who have shared HRSV genomes on the public databases NCBI GenBank, ENA, DDBJ, and
357 GISAID EpiRSV. We express our gratitude for the valuable comments provided by researchers and
358 public health scientists during the initial stages of the RSV Genotyping Consensus Consortium's
359 work.

360

361 **Disclaimers**

362 RAN consults for Moderna on matter in virus evolution. NW has received grant funding from the Bill
363 and Melinda Gates Foundation and Sanofi. The authors received no financial support for the research,
364 authorship, or publication of this article.

365

366 **Bibliography**

- 367 1. GSK. European Commission authorises GSK's Arexvy, the first respiratory syncytial virus (RSV)
368 vaccine for older adults | GSK. 2023. Available from: [https://www.gsk.com/en-gb/media/press-
369 releases/european-commission-authorises-gsk-s-arexvy-the-first-respiratory-syncytial-virus-rsv-
370 vaccine-for-older-adults/](https://www.gsk.com/en-gb/media/press-releases/european-commission-authorises-gsk-s-arexvy-the-first-respiratory-syncytial-virus-rsv-vaccine-for-older-adults/)
- 371 2. Pfizer. U.S. FDA Approves ABRYSVO, Pfizer's Vaccine for the Prevention of Respiratory
372 Syncytial Virus (RSV) in Infants Through Active Immunization of Pregnant Individuals 32-36
373 Weeks of Gestational Age. 2023. Available from: [https://www.pfizer.com/news/press-
374 release/press-release-detail/us-fda-approves-abrysvotm-pfizers-vaccine-prevention-0](https://www.pfizer.com/news/press-release/press-release-detail/us-fda-approves-abrysvotm-pfizers-vaccine-prevention-0)
- 375 3. AstraZeneca. Nirsevimab unanimously recommended by FDA Advisory Committee for the
376 prevention of RSV lower respiratory tract disease in infants. 2023. Available from:
377 [https://www.astrazeneca.com/media-centre/press-releases/2023/nirsevimab-recommended-for-
378 infant-rsv-protection.html](https://www.astrazeneca.com/media-centre/press-releases/2023/nirsevimab-recommended-for-infant-rsv-protection.html)
- 379 4. Anderson LJ, Hierholzer JC, Tsou C, Hendry RM, Fernie BF, Stone Y, et al. Antigenic
380 characterization of respiratory syncytial virus strains with monoclonal antibodies. *J Infect Dis.*
381 1985 Apr;151(4):626–33.
- 382 5. Tian D, Battles MB, Moin SM, Chen M, Modjarrad K, Kumar A, et al. Structural basis of
383 respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion
384 glycoprotein. *Nat Commun.* 2017 Nov 30;8(1):1877.
- 385 6. Salimi V, Viegas M, Trento A, Agoti CN, Anderson LJ, Avadhanula V, et al. Proposal for
386 Human Respiratory Syncytial Virus Nomenclature below the Species Level. *Emerg Infect Dis.*
387 2021 Jun;27(6):1–9.
- 388 7. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. Circulation patterns of genetically
389 distinct group A and B strains of human respiratory syncytial virus in a community. *J Gen Virol.*
390 1998 Sep;79 (Pt 9):2221–9.

HRSV classification below the subgroup level

- 391 8. Trento A, Galiano M, Videla C, Carballal G, García-Barreno B, Melero JA, et al. Major changes
392 in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60
393 nucleotides. *J Gen Virol.* 2003 Nov;84(Pt 11):3115–20.
- 394 9. Eshaghi A, Duvvuri VR, Lai R, Nadarajah JT, Li A, Patel SN, et al. Genetic Variability of
395 Human Respiratory Syncytial Virus A Strains Circulating in Ontario: A Novel Genotype with a
396 72 Nucleotide G Gene Duplication. *PLoS ONE.* 2012 Mar 28;7(3):e32807.
- 397 10. Venter M, Madhi SA, Tiemessen CT, Schoub BD. Genetic diversity and molecular epidemiology
398 of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new
399 subgroup A and B genotypes. *J Gen Virol.* 2001 Sep;82(Pt 9):2117–24.
- 400 11. Cui G, Zhu R, Qian Y, Deng J, Zhao L, Sun Y, et al. Genetic variation in attachment glycoprotein
401 genes of human respiratory syncytial virus subgroups a and B in children in recent five
402 consecutive years. *PloS One.* 2013;8(9):e75020.
- 403 12. Hirano E, Kobayashi M, Tsukagoshi H, Yoshida LM, Kuroda M, Noda M, et al. Molecular
404 evolution of human respiratory syncytial virus attachment glycoprotein (G) gene of new genotype
405 ON1 and ancestor NA1. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis.* 2014
406 Dec;28:183–91.
- 407 13. Blanc A, Delfraro A, Frabasile S, Arbiza J. Genotypes of respiratory syncytial virus group B
408 identified in Uruguay. *Arch Virol.* 2005 Mar;150(3):603–9.
- 409 14. Dapat IC, Shobugawa Y, Sano Y, Saito R, Sasaki A, Suzuki Y, et al. New genotypes within
410 respiratory syncytial virus group B genotype BA in Niigata, Japan. *J Clin Microbiol.* 2010
411 Sep;48(9):3423–7.
- 412 15. Shobugawa Y, Saito R, Sano Y, Zaraket H, Suzuki Y, Kumaki A, et al. Emerging Genotypes of
413 Human Respiratory Syncytial Virus Subgroup A among Patients in Japan. *J Clin Microbiol.* 2009
414 Aug;47(8):2475–82.
- 415 16. Goya S, Galiano M, Nauwelaers I, Trento A, Openshaw PJ, Mistchenko AS, et al. Toward unified
416 molecular surveillance of RSV: A proposal for genotype definition. *Influenza Other Respir*
417 *Viruses.* 2020;14(3):274–85.
- 418 17. Muñoz-Escalante JC, Comas-García A, Bernal-Silva S, Robles-Espinoza CD, Gómez-Leal G,
419 Noyola DE. Respiratory syncytial virus A genotype classification based on systematic
420 intergenotypic and intragenotypic sequence analysis. *Sci Rep.* 2019 Dec 27;9(1):20097.
- 421 18. Ramaekers K, Rector A, Cuypers L, Lemey P, Keyaerts E, Van Ranst M. Towards a unified
422 classification for human respiratory syncytial virus genotypes. *Virus Evol.* 2020 Jul
423 24;6(2):veaa052.
- 424 19. Chen J, Qiu X, Avadhanula V, Shepard SS, Kim DK, Hixson J, et al. Novel and extendable
425 genotyping system for human respiratory syncytial virus based on whole-genome sequence
426 analysis. *Influenza Other Respir Viruses.* 2022 May;16(3):492–500.
- 427 20. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence
428 alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002 Jul 15;30(14):3059–66.
- 429 21. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets.
430 *Bioinformatics.* 2014 Nov 15;30(22):3276–8.

HRSV classification below the subgroup level

- 431 22. A*STAR Bioinformatics Institute. RSVsurver - Respiratory Syncytial Virus Surveillance Server.
432 Available from: <https://rsvsurver.bii.a-star.edu.sg/>
- 433 23. Minh BQ, Schmidt HA, Chernomor O, Schrenpf D, Woodhams MD, von Haeseler A, et al. IQ-
434 TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol*
435 *Biol Evol.* 2020 May 1;37(5). Available from: <https://doi.org/10.1093/molbev/msaa015>
- 436 24. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the
437 Ultrafast Bootstrap Approximation. *Mol Biol Evol.* 2018 Feb 1;35(2):518–22.
- 438 25. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus*
439 *Evol.* 2018 Jan 1;4(1):vex042.
- 440 26. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, et al. Augur: a bioinformatics
441 toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw.* 2021;6(57):2906.
- 442 27. McBroom J, Schneider A de B, Roemer C, Wolfinger MT, Hinrichs AS, O’Toole AN, et al.
443 Automated Agnostic Designation of Pathogen Lineages. *bioRxiv*; 2023. p. 2023.02.03.527052.
444 Available from: <https://www.biorxiv.org/content/10.1101/2023.02.03.527052v1>
- 445 28. O’Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of
446 epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 2021 Dec
447 1;7(2):veab064.
- 448 29. World Health Organization/World Organisation for Animal Health/Food and Agriculture
449 Organization (WHO/OIE/FAO) H5N1 Evolution Working Group. Revised and updated
450 nomenclature for highly pathogenic avian influenza A (H5N1) viruses. *Influenza Other Respir*
451 *Viruses.* 2014;8(3):384–8.
- 452 30. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters
453 based on sequence data. *AIDS Lond Engl.* 2017 Jun 1;31(9):1211–22.
- 454 31. Streng A, Goettler D, Haerlein M, Lehmann L, Ulrich K, Prifert C, et al. Spread and clinical
455 severity of respiratory syncytial virus A genotype ON1 in Germany, 2011–2017. *BMC Infect Dis.*
456 2019 Jul 12;19(1):613.
- 457 32. Goya S, Lucion MF, Shilts MH, Juárez MDV, Gentile A, Mistchenko AS, et al. Evolutionary
458 dynamics of respiratory syncytial virus in Buenos Aires: Viral diversity, migration, and subgroup
459 replacement. *Virus Evol.* 2023;9(1):vead006.
- 460 33. Liang X, Liu DH, Chen D, Guo L, Yang H, Shi YS, et al. Gradual replacement of all previously
461 circulating respiratory syncytial virus A strain with the novel ON1 genotype in Lanzhou from
462 2010 to 2017. *Medicine (Baltimore).* 2019 May;98(19):e15542.
- 463 34. van Niekerk S, Venter M. Replacement of Previously Circulating Respiratory Syncytial Virus
464 Subtype B Strains with the BA Genotype in South Africa. *J Virol.* 2011 Sep;85(17):8789–97.
- 465 35. Hause AM, Henke DM, Avadhanula V, Shaw CA, Tapia LI, Piedra PA. Sequence variability of
466 the respiratory syncytial virus (RSV) fusion gene among contemporary and historical genotypes
467 of RSV/A and RSV/B. *PLoS ONE.* 2017 Apr 17;12(4):e0175792.
- 468 36. Quick J. nCoV-2019 sequencing protocol v1. 2020 Jan. Available from:
469 <https://www.protocols.click/view/ncov-2019-sequencing-protocol-bbmuik6w>

HRSV classification below the subgroup level

- 470 37. Zhou B, Wentworth DE. Influenza A Virus Molecular Virology Techniques. In: Kawaoka Y,
471 Neumann G, editors. Influenza Virus: Methods and Protocols. Totowa, NJ: Humana Press; 2012.
472 p. 175–92. (Methods in Molecular Biology). Available from: [https://doi.org/10.1007/978-1-](https://doi.org/10.1007/978-1-61779-621-0_11)
473 [61779-621-0_11](https://doi.org/10.1007/978-1-61779-621-0_11)
- 474 38. Dong X, Deng YM, Aziz A, Whitney P, Clark J, Harris P, et al. A simplified, amplicon-based
475 method for whole genome sequencing of human respiratory syncytial viruses. *J Clin Virol Off*
476 *Publ Pan Am Soc Clin Virol*. 2023 Apr;161:105423.
- 477 39. Wang L, Ng TFF, Castro CJ, Marine RL, Magaña LC, Esona M, et al. Next-generation
478 sequencing of human respiratory syncytial virus subgroups A and B genomes. *J Virol Methods*.
479 2022 Jan;299:114335.
- 480 40. Presser LD, van den Akker WMR, Meijer A, for PROMISE investigators. Respiratory Syncytial
481 Virus European Laboratory Network 2022 Survey: Need for Harmonization and Enhanced
482 Molecular Surveillance. *J Infect Dis*. 2023 Aug 14;jjad341.
- 483 41. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation
484 calling and quality control for viral genomes. *J Open Source Softw*. 2021 Nov 30;6(67):3773.
- 485 42. Borges V, Pinheiro M, Pechirra P, Guiomar R, Gomes JP. INSaFLU: an automated open web-
486 based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based
487 surveillance. *Genome Med*. 2018 Jun 29;10(1):46.
- 488 43. Turakhia Y, Thornlow B, Hinrichs AS, De Maio N, Gozashti L, Lanfear R, et al. Ultrafast
489 Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-
490 CoV-2 pandemic. *Nat Genet*. 2021 Jun;53(6):809–16.
- 491 44. Redlberger-Fritz M, Springer DN, Aberle SW, Camp JV, Aberle JH. Respiratory syncytial virus
492 surge in 2022 caused by lineages already present before the COVID-19 pandemic. *J Med Virol*.
493 2023;95(6):e28830.
- 494 45. Goya S, Sereewit J, Pfaller D, Nguyen TV, Bakhsh SAKM, Sobolik EB, et al. Genomic
495 Characterization of Respiratory Syncytial Virus during 2022-23 Outbreak, Washington, USA.
496 *Emerg Infect Dis*. 2023 Apr;29(4):865–8.
- 497 46. Adams G, Moreno GK, Petros BA, Uddin R, Levine Z, Kotzen B, et al. RSV Surge in the USA
498 and Viral Lineages, 2022. *N Engl J Med*. 2023 Apr 6;388(14):1335–7.
- 499 47. Acuña D, Goya S, Nabaes Jodar MS, Grandis E, Mistchenko AS, Viegas M. RSV reemergence in
500 Argentina since the SARS-CoV-2 pandemic. *J Clin Virol*. 2022 Apr 1;149:105126.
- 501 48. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, et al.
502 Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region,
503 January to June 2020. *Eurosurveillance*. 2020 Aug 13;25(32):2001410.
- 504 49. Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature
505 proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020
506 Nov;5(11):1403–7.
- 507 50. van Roekel C, Poukka E, Turunen T, Nohynek H, Presser L, Meijer A, et al. Effectiveness of
508 immunisation products against medically attended respiratory syncytial virus infection: generic
509 protocol for a test-negative case-control study. *J Infect Dis*. 2023 Nov 3;jjad483.

HRSV classification below the subgroup level

511 Table 1. Amino acids defining lineages in HRSV-A.

512 The lineage-defining amino acids within each viral protein are detailed. This description provides the
 513 location of the amino acid using the genome coordinates hRSV/A/England/397/2017 (lineage
 514 A.D.2.2.1, NCBI Virus accession number: PP109421.1, GISAID EpiRSV accession number:
 515 EPI_ISL_412866), followed by the amino acid found in the parental lineage in parentheses. Finally,
 516 the amino acid substitution that drives it to become a lineage is specified. Amino acid changes
 517 detected in not all but $\geq 90\%$ of the lineage sequences are indicated with *. The letter X is used to
 518 represent an amino acid deletion.

Lineage	NS1	NS2	N	P	M	SH	G	F	M2-1	M2-2	L
A.1	105(I)L	38(R)K	216(Y)H	73(V)A*	91(C)S		15(K)R; 38(I)V; 71(L)S; 74(A)T; 101(F)P; 104(L)P; 111(T)I; 121(R)G; 141(I)A; 157(N)S; 215(L)P; 222(P)S; 233(K)E; 250(N)S; 254(G)R; 258(H)L; 286(K)M; 304(Y)S; 310(L)P*; 317(S)P; 321(N)R	8(T)A; 16(A)I; 122(T)A		18(N)T; 44(Q)P; 51(P)L; 80(D)E	81(I)L; 104(T)A; 162(L)S; 232(I)T; 240(N)S; 255(H)N; 445(V)I; 603(N)D; 754(R)G; 821(V)I; 1180(F)L; 1657(Y)H; 1718(T)A; 1731(D)E; 1756(T)A; 1773(T)M; 1778(K)R; 1969(T)M; 2019(V)A
A.2							4(N)T; 107(I)T/N; 121(R)S; 123(K)E; 225(V)A; 244(I)R; 251(N)S; 256(P)L*; 263(F)L; 286(K)E*; 314(S)P*	20(F)L; 124(K)N	120(L)P ; 179(L)S	1(M)T*; 25(R)S*; 54(T)P; 68(T)A	174(D)G; 237(Q)H; 1551(K)R; 1721(I)V; 1723(S)C
A.2.1		7(D)G	64(L)V	73(V)A			57(A)V; 108(T)I*; 161(N)D; 303(V)I; 321(N)D	125(T)N; 518(A)V; 547(L)F	182(S)N	48(I)T	148(D)N*; 1724(N)D
A.2.1.1							117(L)P*; 124(S)P; 146(P)L; 206(P)Q	2(E)D; 13(T)A			104(T)A*; 144(G)E; 257(E)D*; 1757(M)T; 1847(N)S
A.3			84(R)K				4(N)T; 15(K)R; 107(I)T; 111(T)P*; 289(F)L	20(F)L*	120(L)P ; 179(L)S	1(M)T; 54(T)P; 68(T)A; 69(I)T	224(N)S; 237(Q)H/Y; 1551(K)R; 1721(I)V; 1764(K)R
A.3.1		38(R)K					106(E)G*; 111(P)S; 237(N)D; 286(K)E; 316(P)S	105(N)S; 574(S)N	125(I)V*	25(S)N; 39(F)I; 50(S)P; 52(N)D	6(N)S*; 177(H)Q*; 216(N)S; 388(V)I; 575(E)D; 754(R)K; 1180(F)L; 1471(H)N; 1700(N)S*; 1730(N)S; 1745(L)S; 2135(Y)H; 2163(H)N
A.3.1.1							142(Q)L; 208(I)L; 297(N)Y	276(N)S; 540(S)A			103(I)T; 240(N)S; 1725(D)E/G; 2163(N)Y
A.D							232(E)G*; 253(T)K*; 260dup>QEETLHST TSEGYLSPSQVYTT SG				598(Y)H*
A.D.1					43(M)I		142(L)S; 320(T)A*				143(N)D; 179(T)S/L; 1653(I)V; 1661(K)N
A.D.1.1							121(S)G; 273(Y)H			2(T)I	124(V)I; 234(I)V; 970(N)S; 1187(D)G; 1236(T)S
A.D.1.2			104(I)F		25(K)R*	51(H)Q	129(T)I*; 174(S)N				1789(S)F

HRSV classification below the subgroup level

A.D.1.3	9(I)L	3(T)I					100(S)I*; 225(V)A; 263(E)Q; 265(L)P; 270(S)P*; 277(S)P; 280(Y)H; 291(S)P; 297(Y)H; 310(L)P; 316(S)P*; 321(K)X; 753(D)G*				755(D)G*; 2111(N)H
A.D.2							134(K)I; 262(E)K			46(N)S	598(H)Y*; 1731(D)G
A.D.2.1			69(T)A			21(I)V	67(H)N; 107(T)P; 120(P)T; 240(K)R; 280(Y)H; 287(E)D; 294(S)P; 303(V)I; 319(T)S				144(G)E; 610(P)S; 1599(H)R
A.D.2.2							240(R)K; 243(I)S; 303(I)V				1731(D)G; 1950(I)L
A.D.2.2.1							224(E)G; 265(L)I; 274(L)P				171(L)P; 256(K)R
A.D.3			92(T)M*				113(T)I; 131(V)D*; 178(N)G*; 258(H)Q*; 266(H)L*	12(T)I		79(T)A	835(L)M
A.D.3.1							101(F)S; 151(R)H; 225(V)A; 230(P)T*; 250(S)F*; 263(E)K	20(L)F; 276(S)N		48(I)L	
A.D.4			63(I)V/A	47(L)Q			57(A)T*; 206(P)Q; 310(L)P*				1726(K)R
A.D.4.1			122(E)D				90(Y)H; 111(S)F*; 283(S)P; 312(Q)X; 319(T)A	123(K)Q; 384(I)T	9(F)Y; 173(I)V	1(T)M	1617(V)I
A.D.5							206(P)Q; 303(V)A*; 320(T)A*	122(T)A			1723(S)G
A.D.5.1			216(Y)H				57(A)V; 118(T)I*; 209(K)R; 319(T)I	23(A)S*			
A.D.5.2			69(T)I				57(A)V; 209(K)R*	103(A)T*		26(C)Y*	
A.D.5.3		3(T)A; 8(T)A					189(I)T; 225(V)A; 243(I)T; 276(P)Q*; 280(Y)H; 284(G)D*; 314(L)P*	15(L)F; 419(K)E	52(K)R; 173(I)V*	27(R)K; 64(D)E	114(R)K; 179(T)A; 335(V)I; 1051(N)S

519

520

HRSV classification below the subgroup level

521 **Table 2. Amino acids defining lineages in HRSV-B.**

522 The lineage-defining amino acids within each viral protein are detailed. This description provides the
 523 location of the amino acid using the genome coordinates HRSV/B/AUS/VIC-RCH056/2019 (lineage
 524 B.D.4.1.1, NCBI Virus accession number: OP975389.1, GISAID EpiRSV accession number:
 525 EPI_ISL_1653999), followed by the amino acid found in the parental lineage in parentheses. Finally,
 526 the amino acid substitution that drives it to become a lineage is specified. Amino acid changes
 527 detected in not all but $\geq 90\%$ of the lineage sequences are indicated with *. The letter X is used to
 528 represent an amino acid deletion.

Lineage	NS1	NS2	N	P	M	SH	G	F	M2-1	M2-2	L
B.1							103(K)I; 133(T)P; 222(K)D	114(Y)H			148(E)D; 948(N)H; 1177(I)M; 1613(K)N; 2014(A)S
B.2	124(D)N	24(E)D		66(I)T			95(S)P*; 118(I)T; 137(I)T; 150(S)P; 206(I)T; 220(M)T; 221(P)T; 275(S)L*; 283(Y)S; 289(S)P; 298(I)T; 311(X)Q*; 314(X)Q	103(V)A*; 209(K)Q; 312(H)P		41(Y)H; 49(C)R*; 68(T)A	308(I)V; 1742(T)M; 2021(N)D
B.3							4(H)N; 77(Q)S; 101(S)P; 137(I)T; 140(S)P; 150(S)P; 152(S)P; 154(N)I; 206(I)T; 221(P)L; 283(Y)S; 298(I)T; 311(X)Q	103(V)A*; 209(K)Q; 312(H)P; 518(T)A*; 529(T)A			184(T)N; 308(I)V; 1043(I)V; 1547(K)R; 1700(N)D; 1764(K)R; 2021(N)D; 2042(T)I
B.4	45(A)T; 105(M)I*				89(I)T	57(L)Q	4(H)N; 32(R)K; 101(S)P; 109(S)P; 137(I)T; 140(S)P; 143(T)N; 150(S)P; 152(S)P; 205(T)P; 206(I)T; 217(P)L; 221(P)L; 235(L)P; 283(Y)S; 298(I)T; 311(X)Q	103(V)A; 209(K)Q; 312(H)P; 529(T)A			184(T)N; 308(I)V; 1043(I)V; 1250(G)S; 1700(N)D; 1726(S)R; 1764(K)R; 2021(N)D; 2042(T)I; 2065(K)N
B.D	138(N)H		372(A)T*				138(T)S*; 157delKK*; 221(L)P*; 227(T)I*; 239dupERDTSTSQS TVLDTTTSKHT; 285(H)Y*; 311(X)Q*; 314(X)Q*	45(F)L*	142(N) S*		177(H)Y; 1716(M)I*; 1787(A)E*
B.D.1		82(T)S			89(T)I		290(E)G*; 311(Q)X*				1593(N)S
B.D.1.1				63(I)V			75(T)S; 204(P)T; 217(L)P; 227(I)T	19(A)T			59(M)I; 1250(G)S
B.D.2				145(I)V			87(I)T; 178(Q)R*; 239(E)K	23(T)A	181(V) I		1250(G)S; 1588(I)L
B.D.3		4(T)K; 80(K)T*					87(I)T; 217(L)P; 268(I)T; 285(Y)H; 299(P)S; 311(Q)X	123(K)G; 124(N)S	188(T) P	3(K)N; 43(H)Y	1489(A)V; 1588(I)L
B.D.4						49(T)I	198(I)T	103(A)V; 172(L)Q	181(V) I		
B.D.4.1			216(H)Y*				107(T)A; 136(R)T; 252(T)I*; 279(I)T	173(S)L			715(I)V*; 1712(T)A
B.D.4.1.1						64(N) D*	131(A)T*; 137(T)I*; 288(T)I*	191(K)R; 206(I)M; 209(Q)R			1479(V)A*; 1716(I)V*

HRSV classification below the subgroup level

B.D.E.1						100(S)G*; 214(P)S; 221(P)L; 252(I)T; 256(K)N*; 268(I)T; 275(S)P; 285(Y)H	190(S)N; 211(S)N; 389(S)P		27(M)T*; 35(D)N; 49(C)F	570(K)R*; 1759(R)K
B.D.E.2			303(I)V			176(N)S; 269(A)V; 274(T)A*				1707(H)N
B.D.E.3		1(M)V				104(Q)H; 128(H)Y; 135(G)S; 141(T)K; 299(P)S*; 311(X)Q*	19(A)V; 526(M)V		75(H)Y*	1965(V)A*
B.D.E.4				61(N)T		135(G)S*; 138(S)F*; 300(T)A; 303(E)G				164(V)M; 309(V)L; 824(I)M*; 1489(A)T/M; 1732(C)Y

529

530

HRSV classification below the subgroup level

531 **Figure 1. The structure and genome of HRSV.**

532 (A) Schematic of the HRSV virion structure detailing the location of structural proteins, created with
533 BioRender.com. (B) Schematic of the HRSV genome organization with the approximated location of
534 genes highlighted (the exact location slightly differs between subgroups and strains). The location of
535 the second hypervariable region in the G gene, used originally for molecular epidemiology
536 classification, is detailed. The location of the G gene 72-nt duplication in HRSV-A and 60-nt
537 duplication in HRSV-B occurs is denoted with a red arrow.

538

539 **Figure 2. The current HRSV genomics surveillance landscape.**

540 Number of HRSV genomes meeting the inclusion criteria used for the classification (NCBI Virus and
541 GISAID review up to March 11, 2023) according to the year of sample collection and subgroup (A)
542 and their country of sample collection (B).

543

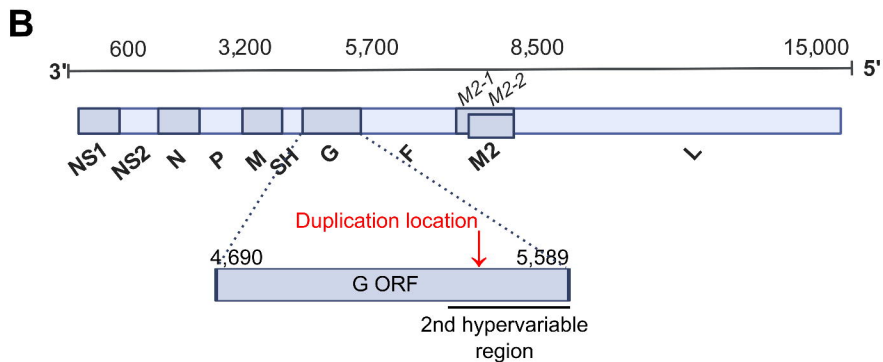
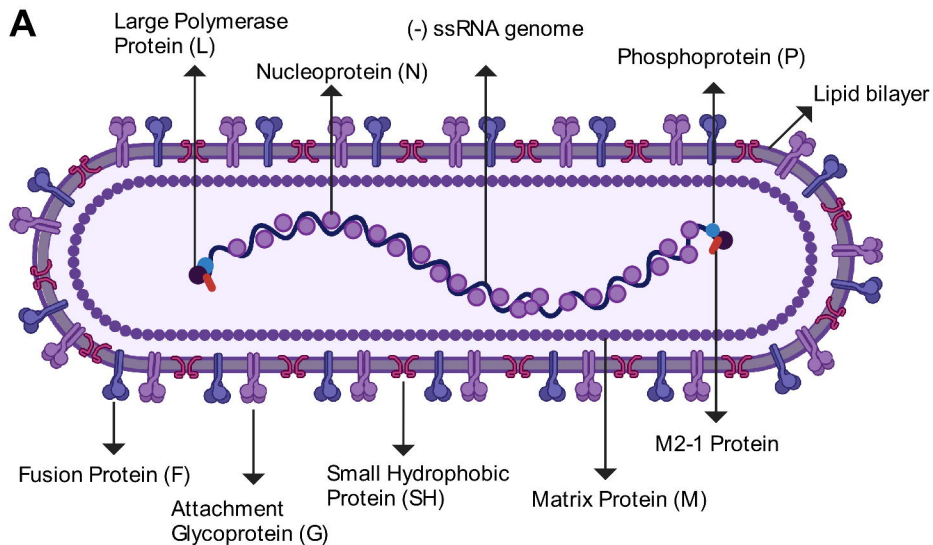
544 **Figure 3. HRSV-A lineages classification and seasonality.**

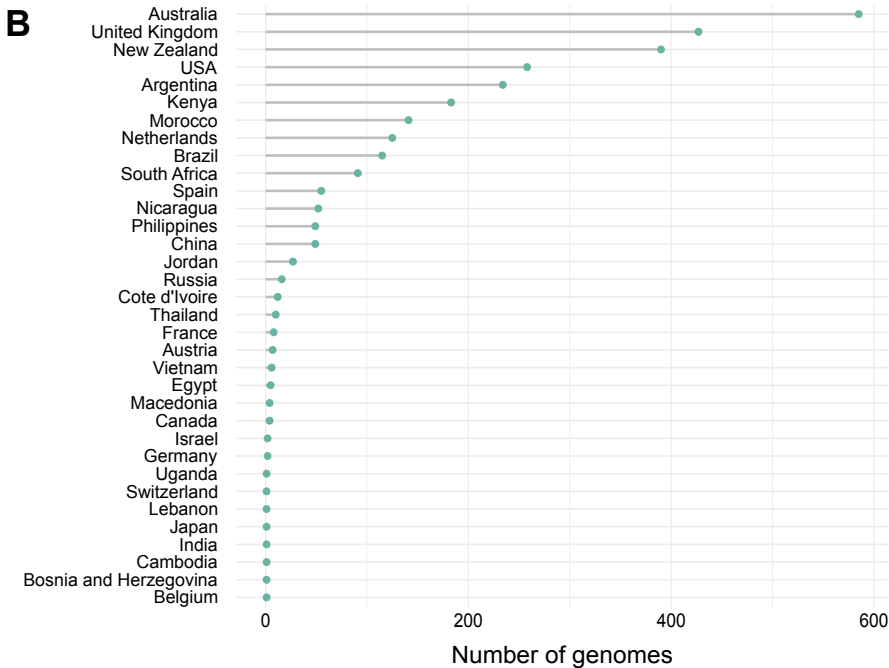
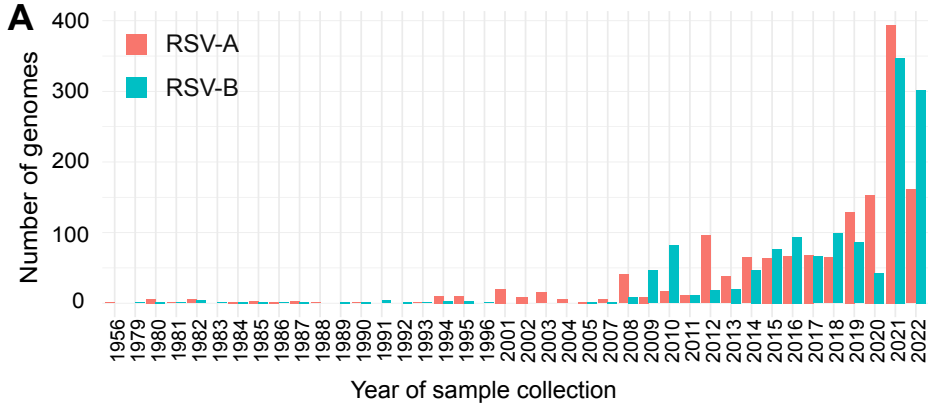
545 (A) The HRSV-A maximum likelihood phylogenetic tree (1,480 sequences) colored according to
546 lineage classification is shown. The A.D lineage defined by the 72-nt duplication in the G gene is
547 denoted with a black star. Scale bar indicates substitutions per site. (B) Simplified scheme of the
548 lineage designation to highlight the presence of nested lineages. The amino acid changes in the F
549 protein are listed next to lineage name and colored according to their location in the fusion protein.
550 (C) Temporal distribution of the HRSV-A lineages including all the genomes available in public
551 databases up to March 2023 (2,744 sequences).

552

553 **Figure 4. HRSV-B lineages classification and seasonality.**

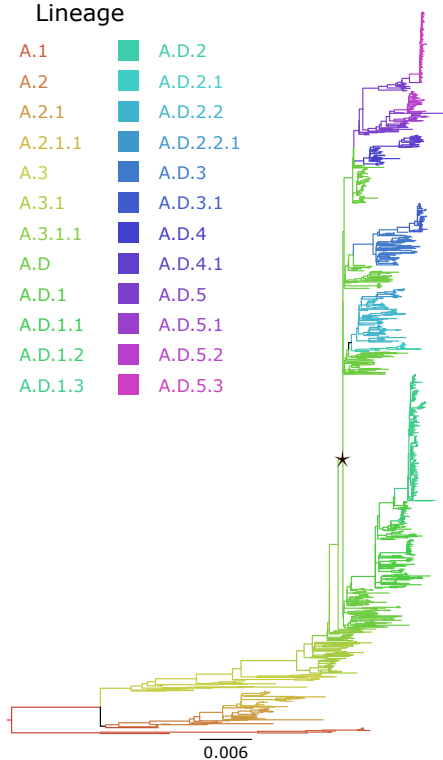
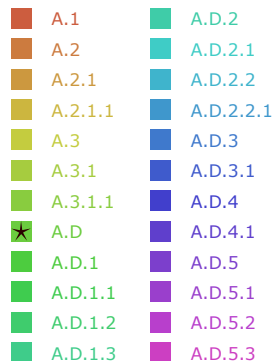
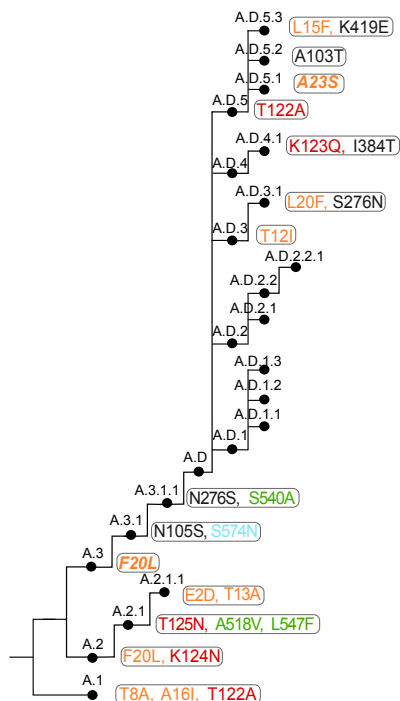
554 (A) The HRSV-B maximum likelihood phylogenetic tree (1,385 sequences) colored according to
555 lineage classification is shown. The B.D lineage defined by the 60-nt duplication in the G gene is
556 denoted with a black star. Scale bar indicates substitutions per site. (B) Simplified scheme of the
557 lineage designation to highlight the presence of nested lineages. The amino acid changes in the F
558 protein are listed next to lineage name and colored according to their location in the fusion protein.
559 (C) Temporal distribution of the HRSV-B lineages including all the genomes available in public
560 databases up to March 2023 (2,443 sequences).





A

Lineage

**B** F protein**C**