Estimating the burden of α-thalassaemia in Thailand using a comprehensive prevalence database for Southeast Asia

Carinna Hockham et al
**Figure 1.** Descriptive maps of the observed allele frequencies in the database. (A) $\alpha^0$-thalassaemia, (B) $\alpha^+$-thalassaemia and (C) $\alpha^{ND}$-thalassaemia. A spatial jitter of up to $0.3^\circ$ latitude and longitude decimal degree coordinates was applied to allow visualisation of spatially duplicated data points. Colour intensity indicates allele frequency, circle size represents the size of the survey size. Surveys that could only be mapped at the national level are indicated by a black star.

DOI: https://doi.org/10.7554/eLife.40580.002
Figure 1—figure supplement 1. A map of the countries included in this study. Here, we defined the Southeast Asian region according to the member states of the Association of Southeast Asian Nations (ASEAN) (http://asean.org/asean/asean-member-states/).

DOI: https://doi.org/10.7554/eLife.40580.003
Figure 1—figure supplement 2. Spatial and temporal distributions of the α-thalassaemia surveys included in the final database. In both panels, the shape of the data points indicates the type of data provided by the survey, the colour indicates whether the survey was found in our online literature search or in local journals, and size represents the sample size of the survey. In (A) a spatial jitter of up to 0.3° latitude and longitude decimal degree coordinates was applied to allow visualisation of spatially duplicated data points.

DOI: https://doi.org/10.7554/eLife.40580.004
Figure 1—figure supplement 3. A map of our current knowledge of the global distribution, gene frequency and genetic diversity of α-thalassemia. Only the most common variants for α⁺-thalassemia (−α²⁺ and −α⁴⁺) and α⁰-thalassemia (−α⁰MED and −α⁰SEA) are shown for each region. The variants that appear in parentheses are those for which the data used to make this map are limited.


DOI: https://doi.org/10.7554/eLife.40580.005
Figure 2. Maps of the mean of, and uncertainty in, the predicted $\alpha$-thalassaemia allele frequencies in Thailand. Panels A to C display the mean of the posterior predictive distribution (PPD) of 100 realisations of the $\alpha$-thalassaemia allele frequency.
Figure 2 continued

geostatistical model. Panels D to F display the 95% credible interval of the PPD. Each row corresponds to a different α-thalassaemia form: α⁺-thalassaemia (A and D; α⁺-thalassaemia (B and E) and αND-thalassaemia (C and F). Figure 2—figure supplement 1 shows the observed data used to construct the models and Figure 2—figure supplement 2 displays the province names for reference.

DOI: https://doi.org/10.7554/eLife.40580.009
Figure 2—figure supplement 1. Maps of the observed allele frequencies used to construct the models and generate the predicted continuous allele frequency maps for Thailand in Figure 2. (A) α-thalassaemia, (B) α-β-thalassaemia and (C) αND-thalassaemia. A variable spatial jitter was applied to allow visualisation of spatially duplicated data points. Colour intensity indicates allele frequency.

DOI: https://doi.org/10.7554/eLife.40580.010
Figure 2—figure supplement 2. A reference map of Thailand provinces. The Bangkok Metropolitan Region, which includes Bangkok City and surrounding provinces, is shaded in red, with Bangkok City shaded a darker red.
DOI: https://doi.org/10.7554/eLife.40580.011
Figure 3. Map showing the proportions of $\alpha^0$, $\alpha^+$- and $\alpha^{ND}$-thalassaemia in Southeast Asia. Three surveys were mapped at the national level (indicated by a white star). The size of the pie charts reflects survey size.

DOI: https://doi.org/10.7554/eLife.40580.014
Figure 4. Map showing the allele frequencies of specific α-thalassaemia variants in Thailand. Given the high number of surveys in northeast Thailand, this region has been magnified. The y-axis scale is the same across all bar charts, ranging from 0 to 1. The variants that were tested for in each survey are indicated above each bar. $\alpha^0$-thalassaemia mutations are shown in red, $\alpha^+-$thalassaemia mutations in blue and $\alpha^{ND}$-thalassaemia mutations in green. Empty spaces along the x-axis indicate an absence of the corresponding mutation in the survey sample. The sample size of the survey is given under each plot. Bar charts are connected to their spatial location by a black line.

DOI: https://doi.org/10.7554/eLife.40580.016
**Figure 5.** Map showing the allele frequencies of specific α-thalassaemia variants in Myanmar, Lao PDR, Cambodia and Vietnam. The y-axis scale is the same across all bar charts, ranging from 0 to 1. The variants that were tested for in each survey are indicated above each bar. α0-thalassaemia mutations are shown in red, α+1-thalassaemia mutations in blue and αND-thalassaemia mutations in green. Empty spaces along the x-axis indicate an absence of the corresponding mutation in the survey sample. The sample size of the survey is given under each plot. Bar charts are connected to their spatial location by a black line. Data points are coloured by country, using the same colour scale as that in Figure 1—figure supplement 1.

DOI: https://doi.org/10.7554/eLife.40580.012
Figure 6. Map showing the allele frequencies of specific α-thalassaemia variants in Malaysia, Singapore and Indonesia. The y-axis scale is the same across all bar charts, ranging from 0 to 1. The variants that were tested for in each survey are indicated above each bar. α0-thalassaemia mutations are shown in red, α+ thalassaemia mutations in blue and αND-thalassaemia mutations in green. Empty spaces along the x-axis indicate an absence of the corresponding mutation in the survey sample. The sample size of the survey is given under each plot. Bar charts are connected to their spatial location by a black line. Data points are coloured by country, using the same colour scale as that in Figure 1—figure supplement 1.

DOI: https://doi.org/10.7554/eLife.40580.018
Figure 7. A schematic overview of the methodology used in this study and a breakdown of the data types analysed. Pink diamonds indicate the database and input data; green boxes denote model processes and data visualisation steps; blue rods represent study outputs.

DOI: https://doi.org/10.7554/eLife.40580.020