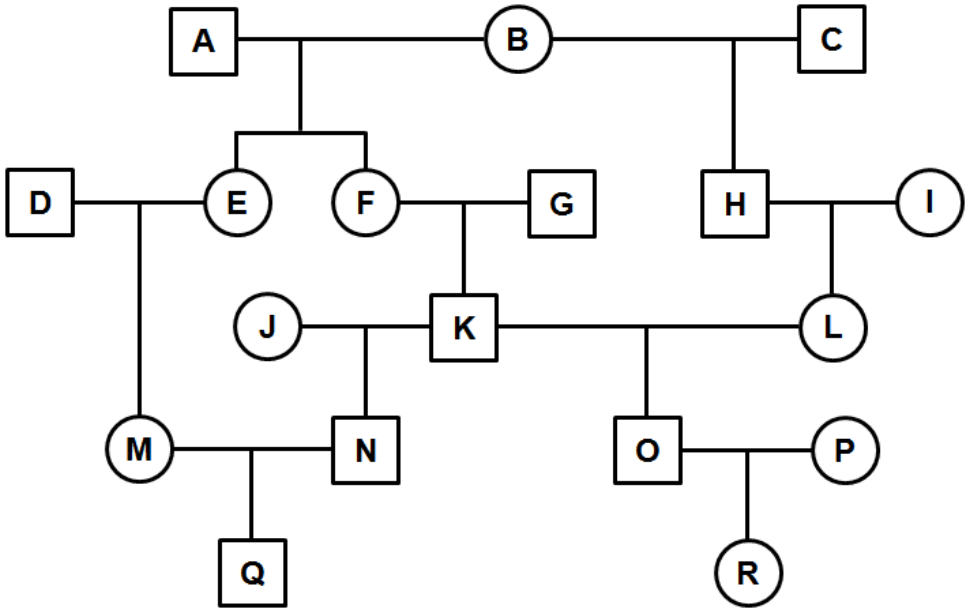
Pedigree-based genetic relatedness

The Genetic covariance between two individuals can be computed using the pedigree information. For individuals A and B, a given pair in a pedigree, the genetic covariance is computed as $r(A,B) = 2\times coancestry(A,B)$ where the *coancestry* between A and B is calculated referring to the method presented by Falconer and Mackay in 1996 (Falconer and Mackay 1996): $coancestry(A,B) = \Sigma_p(1/2)^{n(p)}\times(1 + I_{Common\ Ancestor})$ where p is the number of paths in the pedigree linking A and B, $n(p)$ the number of individuals (including A and B) for each path p and I_X is the *inbreeding* coefficient of X also equal to the *coancestry* between the two parents of X, I_X is set to 0 if X is a founder.

Illustration: Consider, as an example, the pedigree below containing 18 individuals named {A, B, …, R} for the calculation of genetic covariance's.



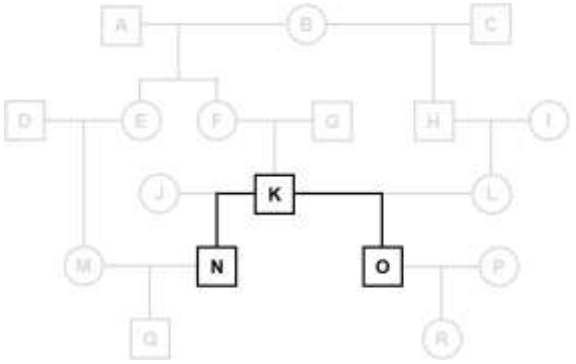
Pedigree structure.

The genetic relatedness between individuals N and O is equal to 0.266. This value is calculated as followed:
The number of paths linking N and O from the pedigree structure above is $p = 2$.
As illustrated below:

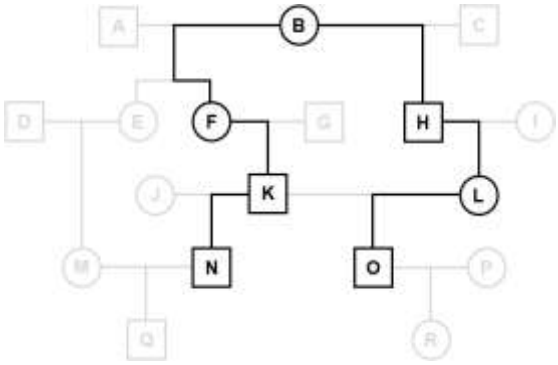
- **Path 1** contains $n(1) = 3$ individuals {N, K, O} with K as the common ancestor. Inbreeding coefficient of K, I_K , is the *coancestry* between the two parents of K (F and G) and is null because F and G are not genetically linked.
- **Path 2** contains $n(2) = 7$ individuals {N, K, F, B, H, L, O} with B as the common ancestor. Inbreeding coefficient of B, I_B , is null because B is a founder.

Therefore, genetic relatedness between individuals N and O is:

$= 2\times(0.5^{n(1)}\times(1+I_K) + 0.5^{n(2)}\times(1+ I_B))$
$= 2\times(0.5^{3}\times(1+0) + 0.5^{7}\times(1+0)) = 0.266$



Path 1 linking N and O.



Path 2 linking N and O.

Defining an equivalent model design where individual effects are independent using the genetic relatedness matrix:

Let us rename $Y^* = l(\mu)$. Y^* can be consider as a linearization of the phenotype through the link function l. The expected mean of Y^* and the variance of Y^* are:

(i) $E(Y^*) = E(X\beta + Z\gamma + \varepsilon)$
$= E(X\beta) + E(Z\gamma) + E(\varepsilon) = X \times E(\beta) + Z \times E(\gamma) + E(\varepsilon)$
$= X\beta$ (asymptomatically).

(ii) $Var(Y^*) = Var(X\beta + Z\gamma + \varepsilon)$
$= Var(Z\gamma + \varepsilon)$ (as $X\beta$ is the fixed part, thus has variance equal to 0)
$= Var(Z\gamma) + Var(\varepsilon)$ (as γ and ε are independent)
$= Z \times Var(\gamma) \times Z^T + Var(\varepsilon)$ (Z^T is the transpose of Z)
$= Z(A\sigma_g^2)Z^T + I\sigma_r^2$
$= ZAZ^T\sigma_g^2 + I\sigma_r^2$

If individuals were independent, i.e. $A = I_N$, variance of Y^* could be expressed as $ZZ^T\sigma_g^2 + I\sigma_r^2$. However, using linear algebra theory by the method "Cholesky decomposition of a matrix", we can show that there is an equivalent expression of the variance of Y^* corresponding to the modeling of data from independent individuals, having γ^* as an equivalent vector of random effects and Z^* an equivalent design matrix relating γ^* to Y^* so that:
$Var(Y^*) = Z^*(I\sigma_g^2)Z^{*T} + I\sigma_r^2$. $I\sigma_g^2$ is then the covariance matrix of the equivalent independent random individual effects γ^*.

Theorem: *Cholesky decomposition of a matrix*
If A is a symmetric positive-definite matrix, there is a triangular matrix L so that A can be written as $A = LL^T$. L can be seen as the "square root" of the matrix A.
Note that the genetic relatedness matrix A computed using the pedigree information (Falconer and Mackay 1996) is a positive-definite matrix, unless identical twins are in the pedigree in which case it would be positive semi-definite.

Equivalent model with independent random effects: We set $A = LL^T$ then:
$Var(Y^*) = Z(A\sigma_g^2)Z^T + I\sigma_r^2$
$= Z(LL^T\sigma_g^2)Z^T + I\sigma_r^2$

$= ZLL^{T}Z^{T}\sigma_g^2 + I\sigma_r^2$
$= (ZL)(ZL)^{T}\sigma_g^2 + I\sigma_r^2$
$= (Z^*)(Z^*)^{T}\sigma_g^2 + I\sigma_r^2$ (where we set $Z^* = ZL$)

Then, if we define $\gamma^* = L^{-1}\gamma$, we can rewrite the model as:

$Y^* = X\beta + Z^*\gamma^* + \varepsilon$ (because $Z\gamma = Z(LL^{-1})\gamma = (ZL)(L^{-1}\gamma) = Z^*\gamma^*$),

and the γ_i^* are independent, in other terms $Var(\gamma^*) = I\sigma_g^2$, as demonstrated below:

We assumed that $\gamma \sim N(0, A\sigma_g^2)$. Then $\gamma^* = L^{-1}\gamma$ is also distributed as a multivariate Normal with mean $E(\gamma^*) = L^{-1}E(\gamma) = L^{-1}\times 0 = 0$ and variance:

$Var(\gamma^*) = (L^{-1})\times Var(\gamma)\times(L^{-1})^{T}$
$= (L^{-1})\times A\sigma_g^2\times(L^{-1})^{T} = (L^{-1})LL^{T}(L^{-1})^{T}\sigma_g^2$
$= (L^{-1}L)(L^{-1}L)^{T}\sigma_g^2$
$= I\sigma_g^2$

The random effects are now independent and then the classical mixed model assuming independence between levels (here individuals) is applied, and the estimate of fixed effects obtained are fine, i.e. corrected for genetic relationships.

References

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. 4th Edn. London: Longman.

Supplementary Tables

Table S1 Number of person-trimesters contributed by number of children by age class and the number who had severe/moderate allergy symptoms, for whom malaria data were also available. AS – Asthma, AD – Atopic dermatitis, RC – Rhinoconjunctivitis. Shown also are the numbers of these individuals suffering from two or all three allergy conditions.

Age group	N° person-trimesters	N° people	AS	AD	RC	AS+AD	AS+RC	AD+RC	AS+AD+RC
]1	7	6	1	2	2	0	1	0	0
]2	21	9	0	1	3	0	0	0	0
]3	48	11	1	1	2	0	0	1	0
]4	119	12	1	2	3	0	0	1	0
]5	102	11	3	4	3	2	1	2	1
]6	125	11	1	1	0	0	0	0	0
]7	303	11	1	2	1	1	0	0	0
]8	340	12	1	1	1	1	0	0	0
]9	362	10	2	0	1	0	1	0	0
]10	610	17	1	0	3	0	0	0	0
]11	77	4	2	1	0	0	0	0	0
]12	484	16	3	0	3	0	1	0	0
]13	390	10	1	0	0	0	0	0	0
]14	105	3	0	0	1	0	0	0	0
Total	3093	143	18	15	23	4	4	4	1

Table S2 Summary of total number of person-trimesters with non-malaria and symptomatic *P. falciparum* clinical presentations and total number of non-malaria episodes according to age class. Given are the number of people contributing to each type of presentation.

	Age group (years)	
	<3·5	**≥3·5**
Total person-trimesters	1283	1810
People	126	113
Total *P. falciparum* symptomatic trimesters	963	1102
People	114	108
Total non-malaria episodes	754	1114
People	123	109

Table S3 Effect of changing age threshold on impact of allergy on the risk of clinical malaria and concomitant parasite density. Given are Odds Ratio with 95% confidence intervals, for clinical malaria episodes and the beta coefficient and standard error for parasite density. Corresponding P values are also given. Values are from the nested GLMM analyses.

A. Malaria episodes							**B. Parasite density**				
Age cut-off (years)	OR	95% CI	P value	OR	95% CI	P value	Age cut-off	beta coeff (se)	P value	beta coeff (se)	P value
Atopy	above threshold			below threshold			Atopy	above threshold		below threshold	
1.5	1.80	1.25-2.59	1.7x10-3	1.57	0.85-2.89	0.15	1.5	0.70 (0.27)	9.2x10-3	0.54 (0.35)	0.12
2.5	2.00	1.39-2.88	2.0x10-4	1.23	0.76-1.99	0.40	2.5	0.79 (0.26)	2.6x10-3	0.35 (0.29)	0.23
3.5	2.02	1.39-2.93	2.1x10-4	1.38	0.92-2.08	0.12	3.5	0.85 (0.26)	9.5x10-4	0.37 (0.26)	0.15
4.5	2.10	1.42-3.10	1.6x10-4	1.41	0.98-2.04	0.063	4.5	0.87 (0.25)	6.9x10-4	0.40 (0.23)	0.09
5.5	1.64	1.07-2.52	0.02	1.67	1.17-2.37	0.004	5.5	0.73 (0.27)	7.4x10-3	0.48 (0.22)	3.4x10-3
Asthma							Asthma				
1.5	1.98	1.29-3.03	1.8x10-3	1.46	0.69-3.19	0.34	1.5	0.66 (0.31)	0.03	0.30 (0.44)	0.48
2.5	2.30	1.49-3.55	1.6x10-4	1.15	0.63-2.09	0.65	2.5	0.78 (0.30)	0.01	0.26 (0.36)	0.48
3.5	2.33	1.50-3.61	1.5x10-4	1.50	0.90-2.50	0.12	3.5	0.82 (0.30)	6.2x10-3	0.43 (0.31)	0.17
4.5	2.30	1.48-3.59	2.4x10-4	1.76	1.11-2.80	0.017	4.5	0.81 (0.29)	5.8x10-3	0.56 (0.28)	0.049
5.5	1.98	1.22-3.22	0.006	2.06	1.33-3.18	0.0011	5.5	0.72 (0.31)	0.02	0.62 (0.27)	0.02
Atopic Dermatitis							Atopic Dermatitis				
1.5	2.05	1.18-3.56	0.01	0.91	0.42-1.97	0.80	1.5	0.80 (0.37)	0.03	0.72 (0.46)	0.12
2.5	2.49	1.36-4.57	3.1x10-3	0.82	0.44-1.53	0.53	2.5	0.77 (0.38)	0.044	0.52 (0.39)	0.19
3.5	3.15	1.56-6.33	1.3x10-3	0.84	0.49-1.46	0.54	3.5	0.99 (0.40)	0.014	0.28 (0.35)	0.42
4.5	3.79	1.61-8.92	2.3x10-3	0.94	0.57-1.57	0.82	4.5	0.98 (0.47)	0.036	0.29 (0.32)	0.37
5.5	1.33	0.47-3.77	0.59	1.19	0.73-1.96	0.49	5.5	0.26 (0.61)	0.67	0.38 (0.31)	0.22
Rhinoconjunctivitis							Rhinoconjunctivitis				
1.5	1.04	0.66-1.62	0.88	1.01	0.51-2.01	0.98	1.5	0.36 (0.32)	0.27	0.18 (0.41)	0.66
2.5	1.01	0.64-1.61	0.96	0.96	0.55-1.68	0.89	2.5	0.28 (0.33)	0.40	0.25 (0.35)	0.48
3.5	0.95	0.60-1.52	0.83	1.05	0.64-1.72	0.85	3.5	0.31 (0.32)	0.33	0.19 (0.31)	0.54
4.5	0.87	0.54-1.42	0.59	1.06	0.68-1.66	0.79	4.5	0.20 (0.32)	0.53	0.22 (0.28)	0.44
5.5	0.81	0.48-1.36	0.43	1.07	0.70-1.64	0.74	5.5	0.10 (0.33)	0.75	0.23 (0.27)	0.39

Table S4 Frequency of non-malaria episodes (number of days of presence divided by number of non-malaria episodes) according to allergic status and age group. The *P* value is that from the GLMM analyses of the effect of allergic status by age group on the number of non-malaria episodes per person-trimester.

Allergic condition	Allergic status (No/Yes)	Age group (years)		*P* value
		<3·5	>3·5	
Atopy	N	78·2	85·9	0·105
	Y	87·2	102·6	
Asthma	N	79·6	87·3	0·319
	Y	82·5	100·2	
Atopic dermatitis	N	80·9	88·2	0·323
	Y	73·4	101·9	
Rhinoconjunctivitis	N	77·9	88·3	0·167
	Y	94·9	91·8	

Figure S1. Incidence of clinical cases per 100 person-trimesters in children under 15 years of age.

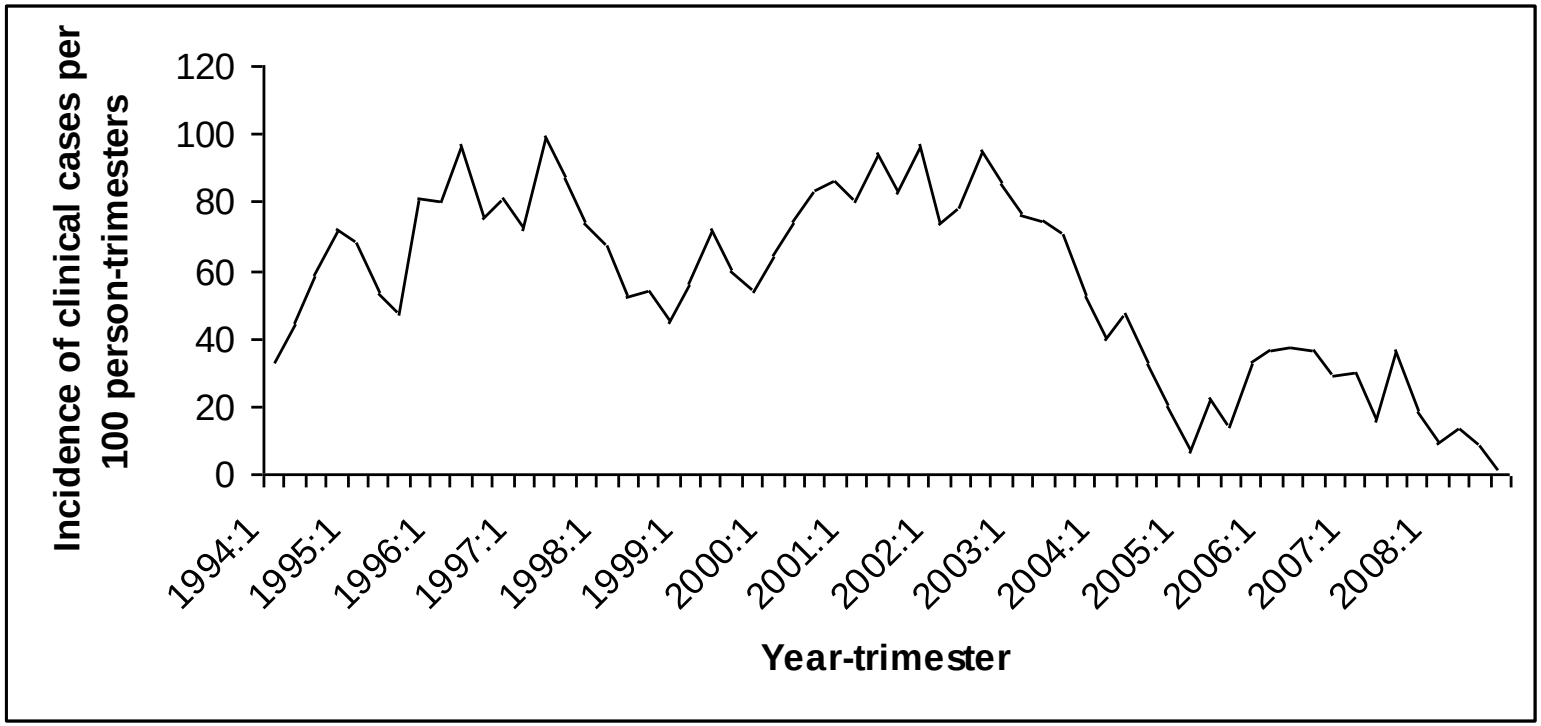


Figure S2. Cumulative incidence of clinical cases according to allergy class predicted by the statistical model.

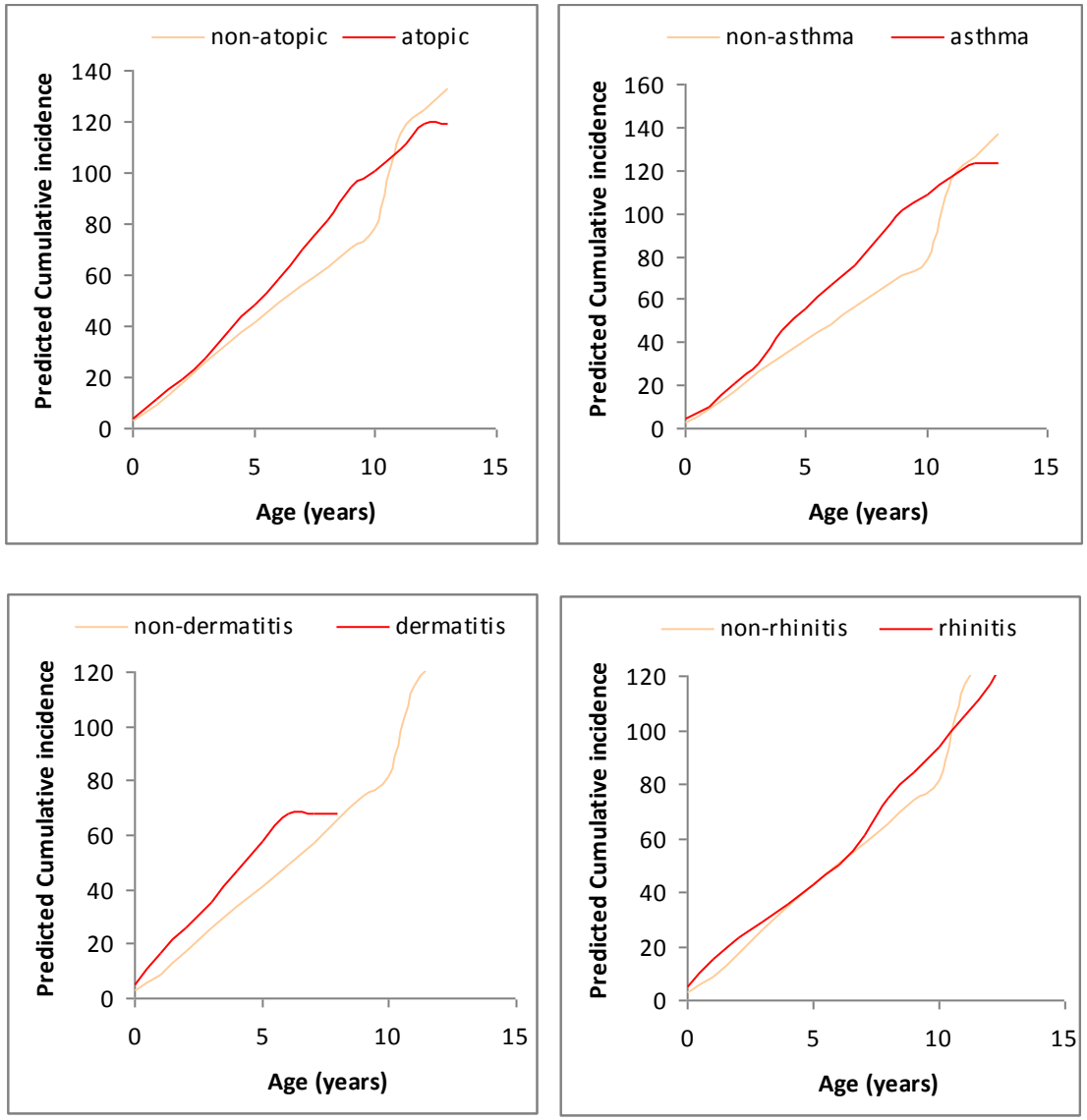
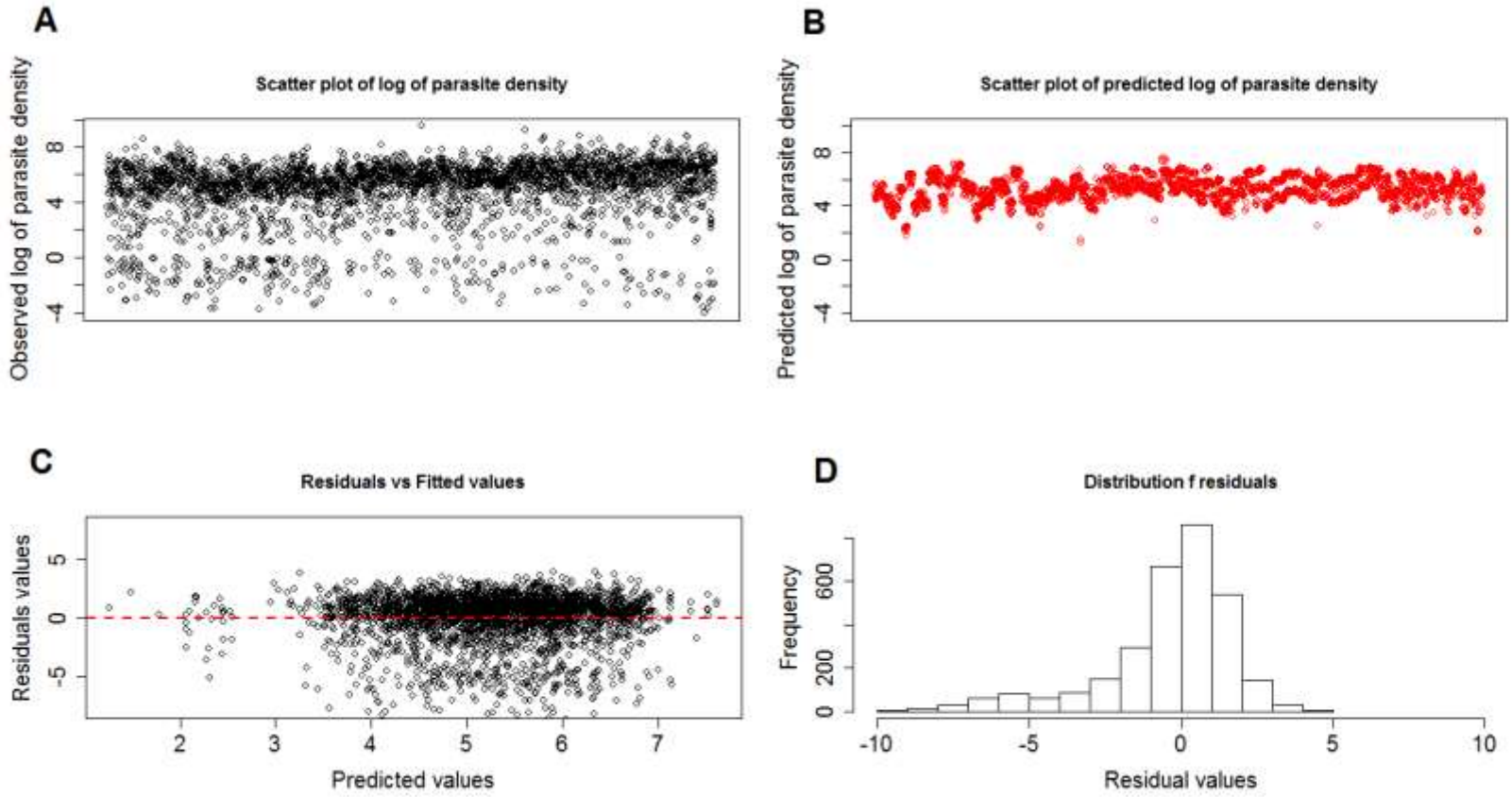


Figure S3. Graphical control model for parasite density

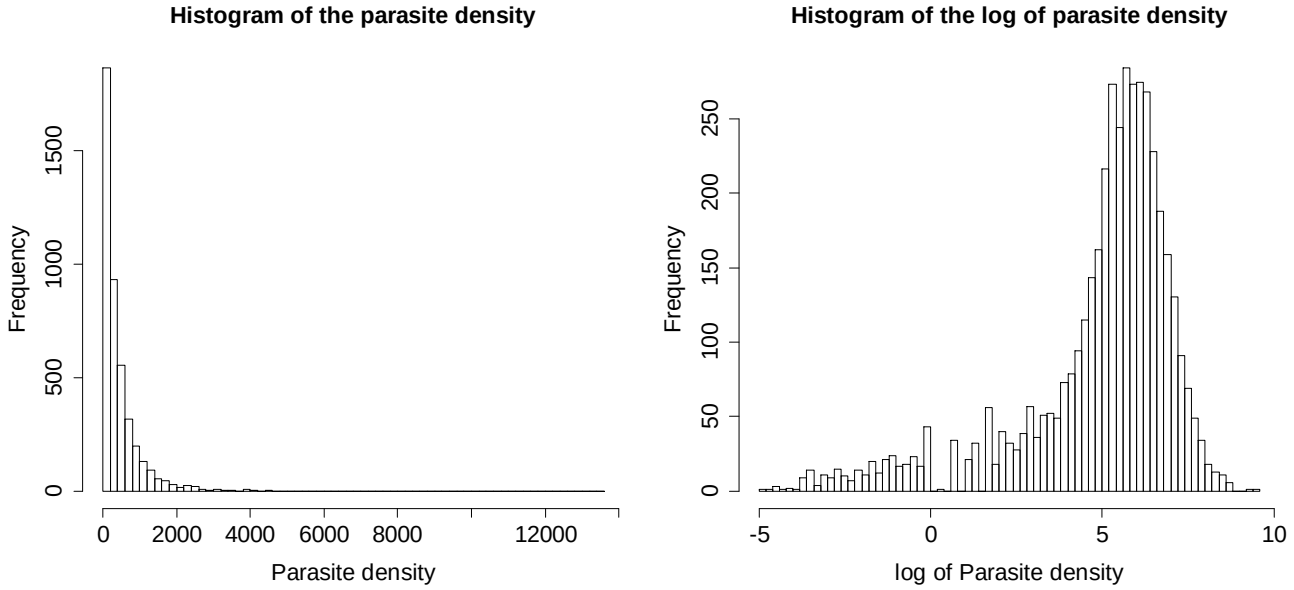
These figures provide a graphical checking of model goodness of fit. Figure A is the scatter plot of the natural logarithm of the observed parasite density and is compared to Figure B, which is the scatter plot of the natural logarithm of the predicted parasite density by the model; on both figures A and B the y-axes give the values for the log of the parasite density. Figure C shows the distribution of the residuals with the predicted values and Figure D is the histogram of the residuals; both figures C and D show the residuals normally distributed around zero.



Analysis using box-cox transformation and probit normalization

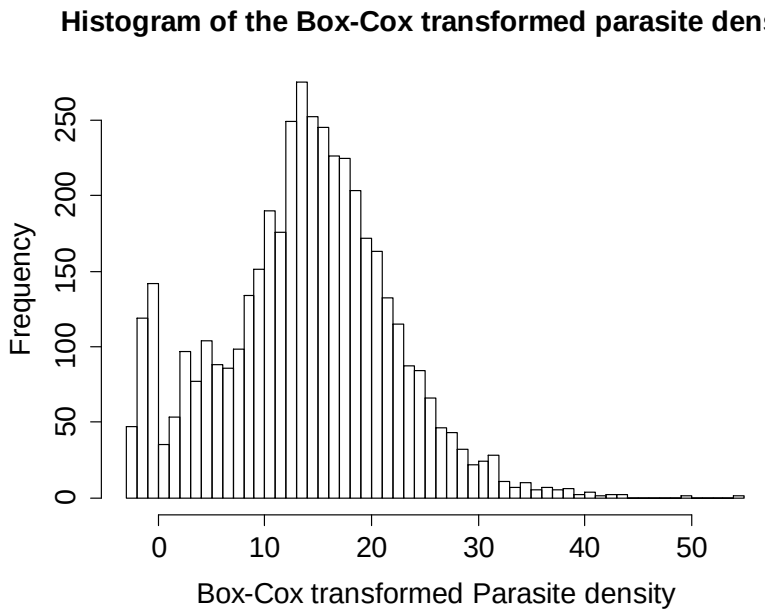
The model we fitted on the parasite density ("*pf_density*") has used as outcome variable the natural logarithm of *pf_density* (equivalent to a Box-Cox for which the parameter is null). As shown on Figure S4 the distribution of log(*pf_density*) is not perfectly normal, it is left-skewed.

Figure S4. Histogram of *pf_density* and log(*pf_density*)



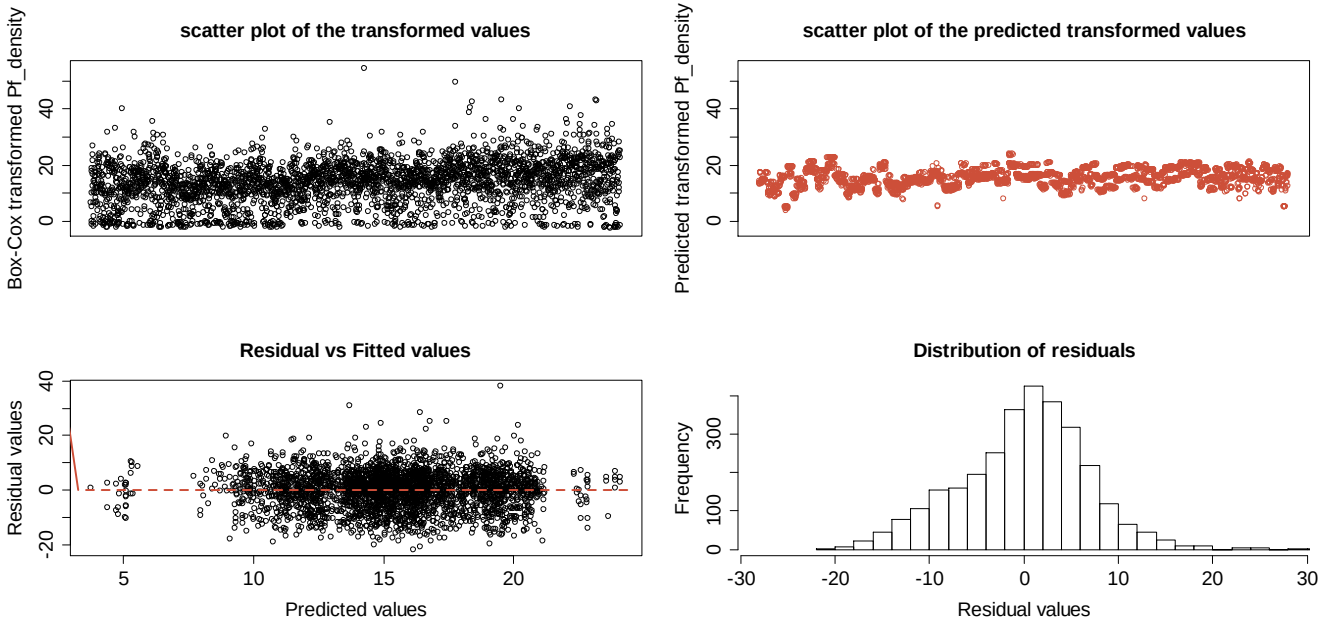
We add here the case for a Box-Cox transformation of the parasite density where the parameter is $\lambda = 0.3$, this parameter value was obtained as optimal using the R- function named "boxcox" from the "MASS" library. Then the Box-Cox transformation of the parasite density is $y = (pf\_density^{0.3} - 1)/0.3$ having the distribution shown on Figure S5 below.

Figure S5. Histogram of the Box-Cox transformation of pf_density using a λ parameter of 0.3



With this Box-Cox transformed parasitemia as outcome variable, our results are maintained. Note that this distribution is not "perfectly" normal. However, the corresponding graphical control of the model adequation presented on Figure S6 below shows residuals more close to the normal distribution than those for log(*pf_density*) as outcome.

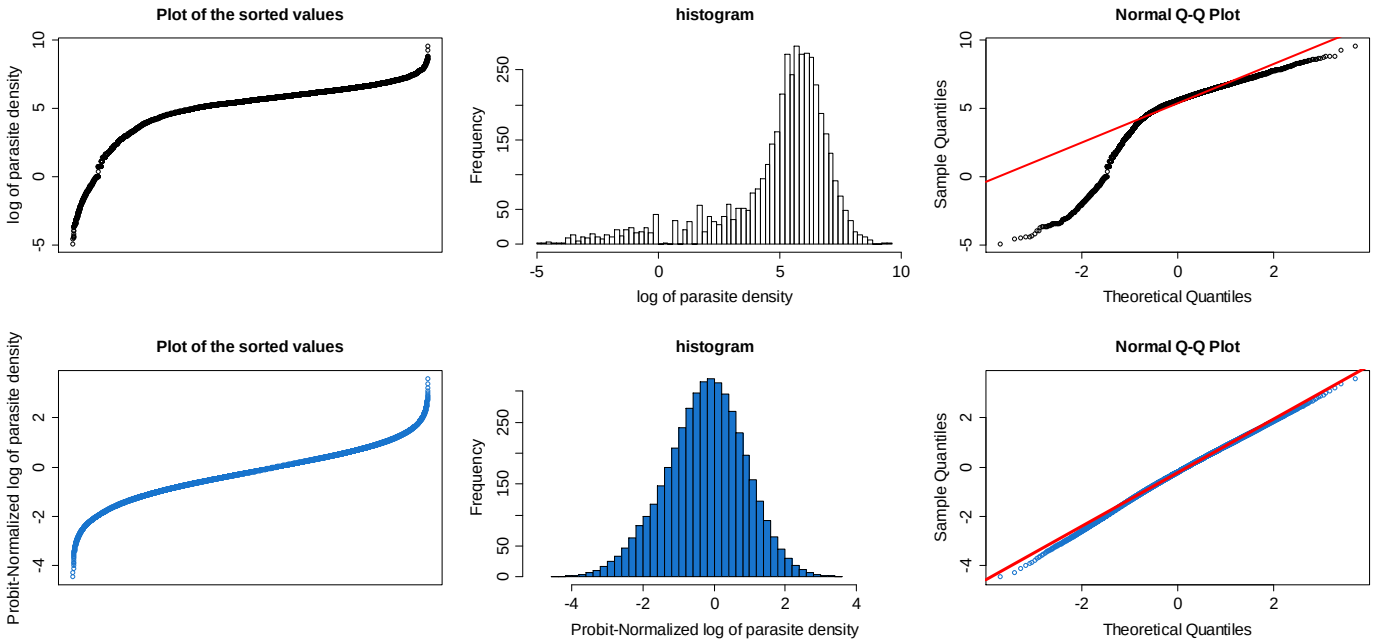
Figure S6. Graphical control of the model adequation for *y* = Box-Cox(*pf_density*, λ = 0.3)



Although using a mixed model approach based on an extreme value distribution would provide a more robust validation of these results, the method we used incorporating pedigree information was developed through an R-package known as "pedigreemm" that allows just for a limited number of distribution laws, which do not include extreme value distributions like the Gumbel or Weibull distributions.

However, we tried the Probit normalization on the log(*pf_density*) to readjust its quantiles to those from a standard normal, and subsequently used the derived standard normal transformation of the log(*pf_density*) as outcome (see Figure S7 below, the three graphs presented in the first row of the graphs panel concern the log(*pf_density*) before Probit normalization and the three in the second row are for after Probit normalization. We can see on the histogram in blue color a good normal distribution of the *y* variable.

Figure S7. Probit normalization of the log(*pf_density*)



The results we obtained after this Probit normalization of the log(*pf_density*) confirmed the same findings. Also, the corresponding graphical control of the model adequation presented on Figure S8 below, shows a good normal distribution of residuals from this model.

Figure S8. Graphical control of the model adequation after Probit normalization of the log(*pf_density*)