

# A statistical framework for genetic association studies of power curves in bird flight

Min Lin<sup>1</sup>, Wei Zhao<sup>1</sup> and Rongling Wu<sup>1\*</sup>

<sup>1</sup>Department of Statistics, University of Florida, Gainesville, FL 32611. USA.

\*Corresponding Author: Rongling Wu, Department of Statistics, University of Florida. Gainesville, FL 32611. USA. Phone: 352-392-3806; Fax: 352-392-8555; Email: rwu@stat.ufl.edu

Submitted: December 5, 2005; Revised: March 28, 2006; Accepted: March 29, 2006.

Indexing terms: QTL; Linkage Disequilibrium; Power Curve; Bird.

---

## ABSTRACT

How the power required for bird flight varies as a function of forward speed can be used to predict the flight style and behavioral strategy of a bird for feeding and migration. A U-shaped curve was observed between the power and flight velocity in many birds, which is consistent to the theoretical prediction by aerodynamic models. In this article, we present a general genetic model for fine mapping of quantitative trait loci (QTL) responsible for power curves in a sample of birds drawn from a natural population. This model is developed within the maximum likelihood context, implemented with the EM algorithm for estimating the population genetic parameters of QTL and the simplex algorithm for estimating the QTL genotype-specific parameters of power curves. Using Monte Carlo simulation derived from empirical observations of power curves in the European starling (*Sturnus vulgaris*), we demonstrate how the underlying QTL for power curves can be detected from molecular markers and how the QTL detected affect the most appropriate flight speeds used to design an optimal migration strategy. The results from our model can be directly integrated into a conceptual framework for understanding flight origin and evolution.

---

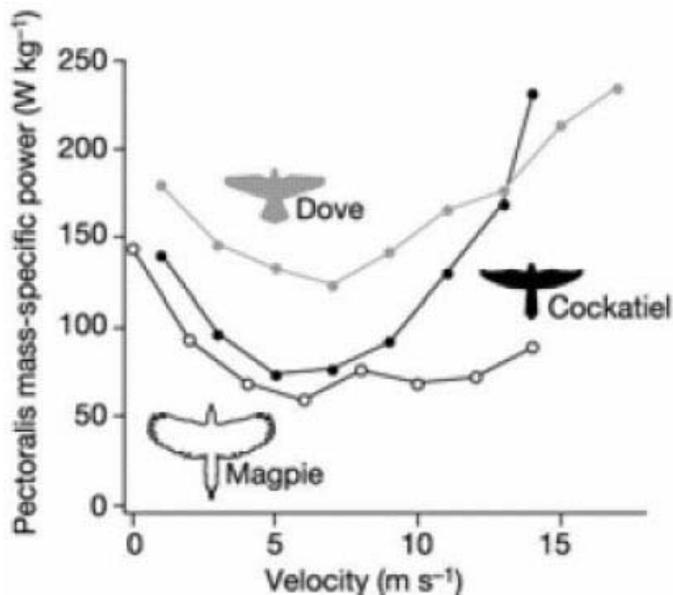
## INTRODUCTION

Our ability to predict the evolutionary trend of avian flight is currently limited by our ability to decipher the genetic architecture of flight performance that involves a series of internal biochemical and physiological processes (1). One promising approach is to integrate two traditionally separated fields, molecular genetics and flight biology, into a unified framework for mapping genetic loci (i.e., quantitative trait loci or QTL) regulating the biological process of bird flight. Flapping flight is among the most energetically expensive activities that vertebrates perform (2, 3). The metabolic cost of flight is a function of forward speeds and can be quantified on the basis of separate rates of energy flow leaving a bird into its environment. The energy budget of a flying bird can be determined by the increase in the kinetic energy of the air caused by the passage of the bird (metabolic

power output) (4) or by the mechanical work performed by the flight muscles at the shoulder point (mechanical power output) (5). Mathematical models have been derived to predict mechanical and metabolic power output (6), but direct measurements of these two outputs can be made possible using three-dimensional kinematic modeling and wind tunnels (7).

The mechanical power required to fly in relation to forward velocity is the sum of three main drag components: induced, parasite and profile drag (Appendix; 8). According to aerodynamic theory applied to flying birds, mechanical power ( $P$ ,  $W$ ) should vary as a function of forward speed ( $V$ ,  $ms^{-1}$ ) in a U-shaped curve (9, 10). But many empirical observations in different birds, such as the black-billed magpie and hummingbirds, suggest that the power curves of bird flight may also be J- or L-shaped (5, 6, 11). Figure 1

presents an excellent example, recently published in *Nature* (7), in which dramatic differences exist in power curve among different bird species. Some physiological and biomechanical explanations are offered about the deviation of the shape of biological power curves from that expected aerodynamic theory (1, 4). It is likely that these different shapes of power curves that have been identified both between and within species (4) include the genetic basis, although little genetic data have been collected.



**Fig. 1: Comparative mass-specific pectoralis power as a function of flight velocity in cockatiels, doves and magpies.** Bird silhouettes are shown to scale, digitized from video. These different power curves can be described by equation (1), with different parameters combinations ( $\alpha, \beta, \gamma$ ). Adapted from Tobalske *et al.* (2003).

In this article, we derive a statistical model for identifying specific genes or quantitative trait loci (QTL) that are responsible for differences in the power curve of bird flight. In principle, this model is a combination of *functional mapping* proposed to map function-valued traits (12) and *linkage disequilibrium mapping* designed to provide high resolution mapping of QTL by making use of recombination events created at a historic time (13). We implement a closed-form solution for the EM algorithm to estimate the population genetic parameters of QTL segregating in a bird population and the simplex algorithm to estimate the curve parameters describing the power of bird in response to different flight speeds. Extensive simulations are performed to investigate the statistical properties of our model. It is expected that our theoretical model can stimulate systematic research into

the genetic basis of ecological and evolutionary processes in birds using genomic mapping approaches.

## MATHEMATICAL MODELLING OF POWER CURVE

Based on physiological and biomechanical principles underlying bird flight, a general parabola has been derived to describe power curve for a flying bird, which is expressed as

$$P = \frac{\alpha}{V} + \beta V^3 + \gamma \quad (1)$$

where the three terms correspond to induced power, parasite power and profile power, respectively (Appendix), and parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  depend on a bird's physical structure, body mass, force and flight style, as demonstrated in Tobalske *et al.* (7). A unique combination ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) of these three parameters determines the shape of power curve in bird flight; different ( $\alpha$ ,  $\beta$ , and  $\gamma$ )-combinations are associated with different curve shapes. The three parameters in power curves can be estimated by fitting equation (1) to observed power data measured at a range of flight speeds based on non-linear regression approaches. But in this study we will develop a mixture model-based likelihood approach to test the existence of a segregating QTL for power curves and estimate these curve parameters for different QTL genotypes.

## POPULATION AND QUANTITATIVE GENETIC MODELS

Suppose there is a natural population for a bird species at Hardy-Weinberg equilibrium, from which a sample of birds (of size  $n$ ) is randomly drawn. In this population, there are  $\nu$  segregating biallelic QTL,  $Q_1, Q_2, \dots, Q_\nu$ , that exert genetic effects individually or epistatically on power curves of flying birds. The genotype frequencies of these  $\nu$  (each with three genotypes) forming  $3^\nu$  genotypes in total, are the products of maternally- and paternally-derived gamete frequencies under the assumption that the population is randomly mating (14). A statistical model will be developed to characterize these QTL based on polymorphic markers, such as single nucleotide polymorphisms (SNPs), genotyped from the avian genome. We attempt to genotype each sampled

bird at  $\varpi$  SNP markers,  $M_1, M_2, \dots, M_{\varpi}$  throughout the entire avian genome to identify all possible QTL based on this newly developed model. The basic assumption for such identification is that there exist significant linkage disequilibria, i.e., non-random associations, among the alleles of QTL and markers. In general, one or more markers can be used to identify a QTL, but it is not recommended to identify multiple QTL based on a single marker.

If all the markers and QTL are located on the same region of a chromosome, there are a total of  $2^{v+\gamma}$  marker-QTL haplotypes. Let  $p_{j_1 j_2 \dots j_{\varpi} k_1 k_2 \dots k_v} (j_1, j_2, \dots, j_{\varpi} = 1, 0$  for two alternative marker alleles  $M$  and  $m$  and  $k_1, k_2, \dots, k_v = 1, 0$  for two alternative QTL alleles  $A$  and  $a$ ) be the haplotype frequency which is expressed as a function of marker and QTL allele frequencies and linkage disequilibria of different orders among these loci (14). To clearly describe this relationship, consider one SNP of alleles  $M$  (in a probability of  $p$ ) and  $m$  (in a probability of  $1 - p$ ), and one QTL of two alleles  $A$  (in a probability of  $q$ ) and  $a$  (in a probability of  $1 - q$ ). With such a one-SNP/one-QTL model, we can, with a relative ease, develop a multiple-SNP/multiple-QTL model. For the one-SNP/one-QTL model, the frequencies of the four marker-QTL haplotypes are expressed as

$$\begin{aligned} p_{11} &= pq + D && \text{for } MA \\ p_{10} &= p(1 - q) - D && \text{for } Ma \\ p_{01} &= (1 - p)q - D && \text{for } mA \\ p_{00} &= (1 - p)(1 - q) + D && \text{for } ma \end{aligned} \quad (2)$$

where  $D$  is the coefficient of linkage disequilibrium between the marker and QTL. Large  $D$  values imply tighter nonrandom associations between the two loci. Similarly, we give a model for specifying the linkage disequilibria among two SNPs,  $M_1$  and  $M_2$ , and one QTL. The frequencies of eight haplotypes constructed by the three loci are expressed as

$$\begin{aligned} p_{111} &= p_1 q p_2 + p_1 D_{A2} + p_2 D_{1A} + q D_{12} + D_{1A2} \text{ for } M_1 A M_2 \\ p_{110} &= p_1 q (1 - p_2) - p_1 D_{A2} + (1 - p_2) D_{1A} - q D_{12} - D_{1A2} \text{ for } M_1 A M_2 \\ p_{011} &= (1 - p_1) q p_2 + (1 - p_1) D_{A2} - p_2 D_{1A} - q D_{12} - D_{1A2} \text{ for } M_1 A M_2 \\ p_{010} &= (1 - p_1) q (1 - p_2) - (1 - p_1) D_{A2} - (1 - p_2) D_{1A} + q D_{12} + D_{1A2} \text{ for } M_1 A M_2 \\ p_{101} &= p_1 q p_2 - p_1 D_{A2} - p_2 D_{1A} + (1 - q) D_{12} - D_{1A2} \text{ for } M_1 A M_2 \\ p_{100} &= p_1 q (1 - p_2) + p_1 D_{A2} - (1 - p_2) D_{1A} - (1 - q) D_{12} + D_{1A2} \text{ for } M_1 A M_2 \\ p_{001} &= (1 - p_1) q p_2 - (1 - p_1) D_{A2} + p_2 D_{1A} - (1 - q) D_{12} + D_{1A2} \text{ for } M_1 A M_2 \\ p_{000} &= (1 - p_1) q (1 - p_2) + (1 - p_1) D_{A2} + (1 - p_2) D_{1A} + (1 - q) D_{12} - D_{1A2} \text{ for } M_1 A M_2 \end{aligned} \quad (3)$$

where  $D_{A2}$ ,  $D_{1A}$  and  $D_{12}$  are the digenic linkage disequilibria between QTL and  $M_2$ ,  $M_1$  and QTL, and  $M_1$  and  $M_2$ , respectively, and  $D_{1A2}$  is the trigenic linkage disequilibrium among the three loci.

Our aim is to estimate the allele frequencies of QTL and their linkage disequilibria with the markers in the population. These population genetic parameters provide important information about the history, structure and dynamics of a bird population (14). Our model will be developed to estimate the haplotype frequencies from which QTL allele frequencies and QTL-marker linkage disequilibria can be estimated by solving equation (1) or (2).

As expected, a QTL forms three possible genotypes. If a QTL affects the power curve, its three genotypes should have different shapes of the curve and therefore different  $(\alpha, \beta, \gamma)$ -combinations. By testing the genotypic difference in these parameter combinations, one can determine whether the QTL affects power curves. However, it is difficult to estimate the curve parameters for different QTL genotypes because the QTL cannot be directly observed. By incorporating QTL genotypes into the mixture model, we derive an algorithm to estimate these curve parameters for each of the three QTL genotypes. Hypothesis tests will provide a basis for claiming genotypic differences in the estimated parameters.

## STATISTICAL METHODS

### Mixture model

The basic statistical model for mapping QTL is a mixture model (14), in which each observation  $y$  is assumed to have arisen from one of  $L$  groups of QTL genotypes, each group being modelled by a density from the parametric family  $f$ . The population density function of  $y$  is expressed as

$$y \sim p(y | \varpi, \phi, \eta) = \varpi f(y; \phi_1, \eta) + \dots + \varpi f(y; \phi_L, \eta) \quad (4)$$

where  $\varpi = (\varpi_1, \dots, \varpi_L)$  are the mixture proportions which are constrained to be non-negative and sum to unity;  $\phi = (\phi_1, \dots, \phi_L)$  are the genotype-specific parameters, with  $\phi_1$  being specific to genotype group 1; and  $\eta$  is a parameter which is common to all genotype groups. The mixture proportions denoted as the frequencies of QTL

genotypes depend on the marker genotypes of a given marker(s) associated with the QTL. The normal density functions associated with different QTL genotypes are expressed in terms of the expected value of each genotype.

In QTL mapping, we use known (observable) marker information to infer unknown (unobservable) QTL genotypes based on their co-segregation in a population. Thus, the mixture proportions of QTL genotypes are determined by the conditional probabilities of QTL genotypes given marker genotypes. These conditional probabilities are expressed in terms of QTL-marker recombination fractions for *linkage mapping* based on controlled crosses (14), or QTL-marker haplotype frequencies for *linkage disequilibrium* (LD) mapping in natural populations (13). For bird species, most of which are still in a wild status, LD mapping based on random samples drawn from a natural population is a better choice for QTL mapping than linkage mapping relied on the availability of well-designed pedigrees. The LD-based strategy has received a surge of interest in QTL mapping because it has proven to be a powerful tool for fine-scale mapping of genes for complex diseases (15, 16). Our model will provide the estimates of these so-called population genetic parameters denoted by  $\Omega_p$ . Wang and Wu (17) proposed a closed form for the EM algorithm to estimate the marker-QTL haplotype frequencies that constitute the conditional probabilities of QTL genotype genotypes given marker genotypes.

In QTL mapping for the power-speed curves in flying birds, each observation  $y$  will be a vector containing the power data at any number (say  $S$ ) of flight speeds. We use vector  $y$  of  $S$  dimensions to denote multivariate phenotypes for each bird. Thus, the phenotypes of each QTL genotype group follow a multivariate normal density

$$f_l = \frac{1}{(2\pi)^{S/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(y - \mu_l)' \Sigma^{-1} (y - \mu_l)\right] \quad (5)$$

where  $\mu$  is the vector of the expected genotypic values of QTL genotype  $l$  for the power trait measured at  $S$  speeds and  $\Sigma$  is the residual variance-covariance matrix of the variables. According to the functional mapping theory put forward in earlier studies (12, 18), the genotypic

values ( $\mu$ ) of the power trait for a QTL genotype at different flight speeds are fit by a biologically meaningful power curve function. Thus, estimating the genotypic values (and therefore additive, dominant and epistatic genetic effects) is equivalent to estimating the mathematical parameters  $(\alpha, \beta, \gamma)$  that specify the shape of power curves. In modelling functional mapping, we usually estimate the structured form of the variance-covariance matrix to increase the model's power rather than estimate every element in the matrix. In Wu *et al.* (18), the first-order autoregressive [AR(1)] model is used, which assumes the stationarity of the residual variance and correlation over different flight speeds. For the AR(1) model, only two parameters, residual variance ( $\sigma^2$ ) and correlation per a unit speed interval ( $\rho$ ), that specify the (co)variance matrix are estimated. Other approaches for structuring the matrix and the advantages of doing so are discussed in (19-21). The parameters contained in the normal distribution are the so-called quantitative genetic parameters denoted by  $\Omega_q$ .

### Computational algorithms

A number of computational algorithms have been developed to estimate unknown genetic values contained in the normal density functions, QTL allele frequencies and QTL-marker linkage disequilibria, arrayed by  $\Omega=(\Omega_p, \Omega_q)$ . They include the maximum likelihood-based EM algorithm (14), Bayesian-based Markov chain Monte Carlo method (22) and optimization techniques used in operations research (23). We have previously derived an efficient computational algorithm for fine mapping of QTL based on the principle of linkage disequilibrium (13). This algorithm offers closed-form solutions for estimating population genetic parameters and represents the most powerful computational tool of its kind by allowing for the estimation of allele frequencies and disequilibria of different orders. Meanwhile, a simplex algorithm (23) is implemented within the EM algorithm to provide the estimates of curve parameters for different QTL genotypes. These curve parameters can be used to estimate various gene actions and gene interactions on bird flight. Our EM-simplex hybrid algorithm allows for the mapping of any possible QTL of gene effects on the differentiation of power curves in bird flight.

## Hypothesis tests

The existence of a QTL that regulates power curves in bird flight can be tested by testing whether different curves each corresponding to a QTL genotype are overlapped. This is equivalent to the null hypothesis ( $H_0$ ) formulated as  $\alpha_1=\alpha_2=\alpha_3$ ,  $\beta_1=\beta_2=\beta_3$  and  $\gamma_1=\gamma_2=\gamma_3$ , and, where the subscripts 2, 1, and 0 denote the QTL genotype  $QQ$ ,  $Qq$  and  $qq$ , respectively. The alternative hypothesis ( $H_1$ ) proposes that these three curves are not overlapped and therefore a QTL exists. The test statistic for testing the hypotheses  $H_0$  vs.  $H_1$  is calculated as the log-likelihood ratio (LR)

$$LR = -2[\ln L(\tilde{\Omega}_p, \tilde{\Omega}_q) - \ln L(\hat{\Omega}_p, \hat{\Omega}_q)] \quad (6)$$

where the tildes and hats denote the MLEs of the unknown parameters under the  $H_0$  and  $H_1$ , respectively. The LR is asymptotically  $\chi^2$ -distributed with 6 degrees of freedom. An empirical approach for determining the critical threshold is based permutation tests. By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of the maximum log-likelihood ratios are calculated, from the distribution of which the critical threshold is determined.

One of the significant strengths of our model is that it allows for the tests of a number of ecologically or evolutionary hypotheses which is not possible for existing genetic mapping models. These hypotheses include whether the QTL detected also exerts an effect on two ecologically important flight speeds, the minimum power speed ( $V_{mp}$ ) and the maximum range speed ( $V_{mr}$ ). These two speeds affect the strategies for bird migration (8).

To test the effect of the power curve QTL on  $V_{mp}$  and  $V_{mr}$ , we pose constraints

$$\frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2} = \frac{\alpha_3}{\beta_3},$$

and

$$\frac{\alpha_2\beta_0 - \alpha_0\beta_2}{\beta_2\gamma_0 - \beta_0\gamma_2} = \frac{\alpha_1\beta_2 - \alpha_2\beta_1}{\beta_1\gamma_2 - \beta_2\gamma_1} = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{\beta_0\gamma_1 - \beta_1\gamma_0}$$

respectively. The corresponding null and alternative

hypotheses for each of these two tests can be formulated, whose LR test statistics are calculated. The determination of critical thresholds for these two tests is based on simulation studies, in which marker and phenotype data under the null hypothesis are simulated repeatedly (e.g., 1,000 times) to estimate the 1%-percentile of the LR distribution.

## MONTE CARLO SIMULATION

To demonstrate the power of our model for functional mapping of power curves, we use a simulated data set to mimic the relationship between the metabolic power and flight speed established in starlings (*Sturnus vulgaris*) detected by Ward *et al.* (4). These authors measured metabolic power at 8 different flight speeds in the European startling *Sturnus vulgaris*. They used a respirometry mask to measure the rates of oxygen consumption and carbon dioxide production during wind tunnel flight. The parameters of power curves fitted to their data were estimated using the least squares approach. We hypothesized different curves within the ranges of these parameters, each curve corresponding to a QTL genotype. Marker genotypes of arbitrarily many loci were simulated for a number of assumed starling birds based on given allele frequencies and linkage disequilibria (Tables 1 and 2), whereas power data at 8 different flight speeds were simulated for each QTL genotype based on the multivariate normal distribution (equation (4)).

We simulated 200 starlings as a random sample from a panmictic population, in which  $\nu$  QTL determine part of inter-individual variation in the shape of power curves. There is the variation among these 200 simulated birds in metabolic power required to fly in a relation to flight speeds due to the effects of the underlying QTL and residual errors. The residual errors followed a multivariate normal distribution with the zero mean vector and the residual (co)variance matrix specified by  $\sigma^2$  and  $\rho$ . The determination of  $\sigma^2$  depends on the size of the heritability of the QTL. In this study, we compare the results from two different heritability levels, 0.1 and 0.4. We simulated the genotypes of the birds at  $\omega$  SNPs, of which some are associated with the QTL underlying power curves, whereas the others display no disequilibria with the QTL. The simulated power curves and SNP genotypes are analyzed by our functional mapping model.

**Table 1: Maximum likelihood estimates of the population genetic parameters describing the three power curves, each corresponding to a QTL, and marker allele frequency, QTL allele frequency and marker-QTL linkage disequilibrium with 8 speeds for a sample size of 200. The numbers in parentheses are the sampling errors of the estimates (One-SNP/one-QTL model).**

Parameters	Genotype	True value	Heritability	
			0.1	0.4
$\alpha$	AA	2.7	2.6386(0.3036)	2.6901(0.0766)
	Aa	2	2.0425(0.2940)	1.9876(0.1473)
	aa	4	4.0720(0.2365)	4.0047(0.0806)
$\beta$	AA	0.0006	0.0006(9.7e-5)	0.0006(2.4e-5)
	Aa	0.0008	0.0008(7.9e-5)	0.0008(4.2e-5)
	aa	0.0012	0.0012(5.8e-5)	0.0012(2.8e-5)
$\gamma$	AA	4.25	4.3492(0.3028)	4.2516(0.0603)
	Aa	5.08	5.0582(0.1497)	5.0772(0.1150)
	aa	4.81	4.8156(0.1183)	4.8081(0.0574)
$\rho$		0.6	0.5922(0.0292)	0.5974(0.0288)
$\sigma^2$		0.95	0.9335(0.0620)	
		0.16		0.1591(0.0224)
$p$		0.6	0.5969(0.0246)	0.6017(0.0231)
$q$		0.6	0.5789(0.0660)	0.5948(0.0264)
$D$		0.08	0.0846(0.0182)	0.0795(0.0161)

For simplicity, we first explain our results from a one-SNP/one-QTL model, with the results compared with those from a two-SNP/one-QTL model. For the one-SNP/one-QTL model, the SNP with allele frequency  $p=0.6$  is associated with the QTL with allele frequency  $q=0.6$ , both loci having a linkage disequilibrium of  $D=0.08$ . Our results suggest that the QTL responsible for power curves can be detected using the SNP in association with the QTL. The estimated LR values for the heritability of 0.1 and 0.4 range 110–255 and 245–487, respectively, strikingly larger the critical threshold, 24.67 ( $P=0.01$ ), calculated from simulation studies. The parabola parameters ( $\alpha, \beta, \gamma$ ) for the power curve of each QTL genotype can be estimated accurately (Table 1), having the estimated values consistent with the hypothesized values under the power curve heritability of 0.1 (Fig. 2). The population genetic parameters of the QTL can be estimated with reasonably high precision using our closed-form solution approach. The means of the estimates from 100 simulations are obtained as  $p=0.597$  and  $q=0.583$  for allele frequencies and  $D=0.085$  for marker-QTL linkage disequilibrium. We compare the power of our model under different heritability levels, disequilibrium levels and sample sizes. The QTL effects as reflected by the differences in curve parameters, QTL allele frequencies and QTL-marker disequilibria can be estimated more precisely for a QTL displaying a greater proportion of the observed variation and a higher QTL-

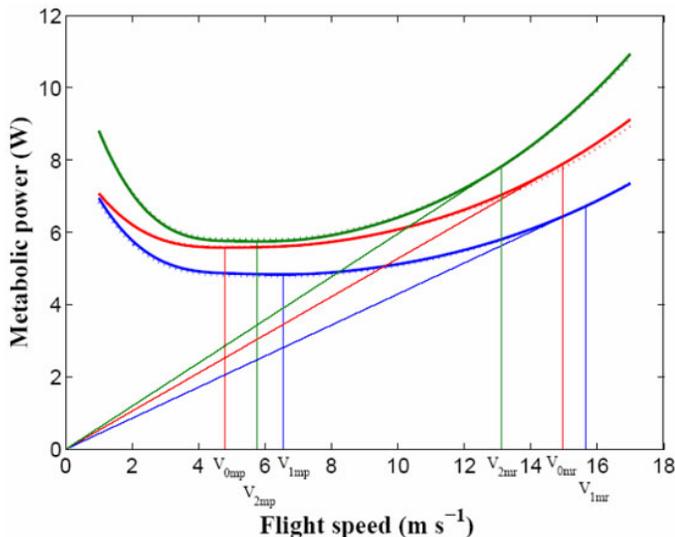
marker disequilibrium than for displaying a smaller proportion and lower disequilibrium (see Table 1 for some of the results). According to our study, a sample of 200 is reliably enough to obtain precise parameter estimation for a modest heritability of power curve (0.1) although increased sample sizes and higher heritabilities can improve the estimation precision.

**Table 2: Maximum likelihood estimates of the quantitative genetic parameters describing the three power curves, each corresponding to a QTL, and marker allele frequency, QTL allele frequency and marker-QTL linkage disequilibrium with 8 speeds for a sample size of 200. The numbers in parentheses are the sampling errors of the estimates (Two-SNP/one-QTL model).**

Parameters	Genotype	True value	Heritability	
			0.1	0.4
$\alpha$	AA	2.7	2.6754(0.3607)	2.6964(0.0910)
	Aa	2	2.0134(0.2031)	2.0035(0.0559)
	aa	4	3.9981(0.2319)	3.9958(0.0630)
$\beta$	AA	0.0006	0.0006(9.8e-5)	0.0006(2.8e-5)
	Aa	0.0008	0.0008(6.0e-5)	0.0008(1.5e-5)
	aa	0.0012	0.0012(6.9e-5)	0.0012(2.0e-5)
$\gamma$	AA	4.25	4.2809(0.2771)	4.2581(0.0543)
	Aa	5.08	5.0658(0.1087)	5.0775(0.0294)
	aa	4.81	4.8166(0.1185)	4.8157(0.0401)
$\rho$		0.6	0.5951(0.0164)	0.5960(0.0139)
$\sigma^2$		0.95	0.9381(0.0352)	
		0.16		0.1568(0.0053)
$p_1$		0.6	0.5982(0.0252)	0.5969(0.0252)
$q$		0.6	0.5971(0.0507)	0.5908(0.0347)
$p_2$		0.6	0.5977(0.0249)	0.5980(0.0250)
$D_{1A}$		0.08	0.0808(0.0204)	0.0792(0.0163)
$D_{A2}$		0.08	0.0794(0.0194)	0.0773(0.0158)
$D_{12}$		0.08	0.0811(0.0160)	0.0815(0.0163)
$D_{1A2}$		0.01	0.0107(0.0109)	0.0106(0.0086)

Our model provides an elegant means for testing the genetic control of ecologically significant or ‘optimal’ flight speeds, such as the minimum power speed ( $V_{mp}$ ) and the maximum range speed ( $V_{mr}$ ) (8).  $V_{mp}$ , obtained from  $dP / dV = 0$ , is associated with a minimum rate of fuel consumption, with no consideration of flight distance, whereas  $V_{mr} (>V_{mp})$ , obtained from  $dP / dV = P / V$  is associated with a minimum energy cost per unit distance flown. At  $V_{mp}$ , birds can be kept airborne for the maximum duration using only the take-off fuel store. Birds minimizing energy cost on migration should fly at  $V_{mr}$ , but the birds should fly slightly faster than  $V_{mr}$  if they are minimizing the overall migration time (8). In the simulation example derived from the starling experiment, we found that the QTL for overall power curves exerts significant effects on the speed for

minimum power ( $LR_{mp} = 101.5$ ,  $P < 0.001$ ) and the speed for maximum flight range ( $LR_{mr} = 102.2$ ,  $P < 0.001$ ). The differences in these two speeds among different QTL genotypes are illustrated in Fig. 2, in which  $V_{mr}$  can be found by drawing a tangent from the origin to the power curve.



**Fig. 2: Estimated power curves (solid) for each of the three QTL genotypes, QQ (green), Qq (blue) and qq (red), in a comparison with the hypothesized curves (dot) used to 200 simulate individual power curves (under the heritability of 0.4).** The consistency between the estimated and hypothesized curves suggests that our model can provide the precise estimation of the genetic control over power curves in flying birds. The differences among the three curves are highly significant ( $LR = 395$ ,  $P < 0.001$ ), suggesting that the assumed QTL plays a pivotal role in shaping the power curve in birds. This power curve QTL is further tested for its genetic effects on two ecologically important flight speeds, the minimum power speed ( $V_{mp}$ ) and the maximum range speed ( $V_{mr}$ ). The three genotype-dependent values for each of these two speeds differ significantly ( $P < 0.001$ ), implying that the detected QTL also affects the two speeds and, thus, the strategy for bird migration.

As expected, when more associated SNPs are used to map a QTL, the accuracy and precision of parameter estimation can be improved, but depending on the level of heritability. The greater improvement occurs when a power curve displays a low heritability. For example, under the curve heritability of 0.1, the MLE of the QTL allele frequency is 0.5971 from the two-SNP/one-QTL model (Table 2), which is closer to the true value 0.6, compared with 0.5789 from the one-SNP/one-QTL model (Table 1). The sampling error of the MLE is also smaller from the two-SNP/one-QTL model (0.0026) than from the one-SNP/one-QTL model (0.0043). The advantage of the two-SNP/one-QTL over one-SNP/one-QTL model becomes more pronounced when the marker and QTL display weak disequilibria and when a small sample size is used (result not shown).

## DISCUSSION

We build up a statistical framework for mapping specific QTL that account for variation in metabolic rates at different flight speeds characterized by a mathematical equation of power curves (equation (1)). We based our modelling on a general power curve -- parabola -- in bird flight constructed from the aerodynamic theory (10) and energetic cost of locomotion (1, 4). The framework embraces two advanced mapping strategies -- linkage disequilibrium analysis with power for fine scale mapping of QTL (13, 15, 16) and functional mapping capable of detecting biologically meaningful QTL (12, 18). As demonstrated by simulation studies, this framework can be well used to detect the QTL that affect bird flight.

Although our idea was described by a simple one-QTL model based on one or two SNPs, its extension to use any number of markers has been derived although such a linkage disequilibrium analysis with high-dimensional SNP data will face computational burden. The number of SNPs used for association studies seems to rely upon the structure and organization of the avian genome. In humans, empirical molecular studies have suggested that the human genome can be broken into a series of discrete haplotype blocks based on different associations of SNPs measured by the coefficients of linkage disequilibrium (24-26). In each haplotype block, consecutive sites are in complete (or nearly complete) linkage disequilibrium with each other and there is limited haplotype diversity due to little (coldspot) inter-site recombination. Adjacent blocks are separated by sites that show evidence of historical recombination (hotspot). Actually, there is a low haplotype diversity within blocks and, therefore, it is possible that a small number of haplotype-tagging SNPs (htSNPs) can be identified for explaining a large portion of haplotype diversity. If a block-like structure is also found in birds, a relatively low number of htSNPs can be well used to detect the underlying QTL for bird flight.

Our model allows for the incorporation of any biochemical and physiological properties related to bird flight (3). Its power is demonstrated by its ability to reliably estimate the QTL for power curves in a flying bird and to test whether the power curve QTL trigger significant effects on the most appropriate flight speeds for the bird's optimal migration strategy. The model can

be extended to include the pleiotropic effect of QTL on both the magnitude and shape of power curves and the physical structure, body mass and flight style of birds (7). Also, a number of statistical issues in power curve mapping remain solved that include optimal structuring of the (co)variance matrix, missing data problems and QTL  $\times$  QTL and QTL  $\times$  environment interactions. These issues will be discussed elsewhere.

Although it is statistically an extension of our previous functional mapping models developed to map growth trajectories in forest trees, our model presented in this study is original in that it provides a unique idea and tool for addressing fundamental ecological and evolutionary problems in bird flight. We anticipate that its publication can stimulate the integration of traditional avian biology and avian genomics to gain new insights into the genetic architecture of any biological processes related to bird flight. Thus, the most appealing and powerful feature of our model is that its deployment in an avian biological research project can shed great light on the origin and evolution of animal flight.

## APPENDIX: MECHANICS OF POWERED BIRD FLIGHT

Power ( $P$ ) required for flight in relation to forward speed ( $V$ ) is obtained from the sum of three main drag components: induced, parasite and profile drag.

### Induced power

A simple approach to calculate induced drag represents the bird as circular disk ( $S_d$ ) with the wing span as diameter, which in aeronautical terms is an actuator disk that produces an induced downwards velocity to the air flowing through it. The mass flow through the wing disk is the volume flow,  $S_d = \mu b^2 / 4 \times$  forward speed ( $V$ )  $\times$  air density ( $\rho$ ), and written as  $S_d V_p$ . By multiplying the mass flow (unit  $\text{kg S}^{-1}$ ) by the speed acquired in the far wake ( $2v_i$ , where  $v_i$  is the induced velocity of the flow around the wing disk) we get the force by which the bird is pushing on the air to support its weight as  $mg = 2v_i S_d V_p$ . This can be rearranged to determine the induced velocity as  $v_i = mg / (2S_d V_p)$ . Power is force  $\times$  speed and the power required to generate the induced velocity, or equivalently to generate a sufficient lift to balance the weight, is the induced power:

$$P_{ind} = \frac{k(mg)^2}{2S_d V_p}$$

where  $k$  is the induced power factor that accounts for deviations from the ideal elliptic lift distribution.

### Parasite power

Parasite drag is the drag of the body isolated from the wings and is calculated as

$$D_{par} = \frac{1}{2} \rho S_b C_{D,par} V^2$$

where  $S_b$  is the body frontal area and  $C_{D,par}$  is a dimensionless drag coefficient. Parasite drag occurs mainly because the body causes an increase of pressure in front of it, which tends to decelerate oncoming flow. A large frontal area causes a relatively high drag, whereas a slim and streamlined body allows oncoming air molecules to flow past more easily resulting in low drag. The parasite power  $P_{par}$  is simply parasite drag  $\times$  forward speed as

$$P_{par} = \frac{1}{2} \rho S_b C_{d,par} V$$

### Profile power

The wings also cause profile drag in addition to the induced drag associated with lift generation. If the wings are held in a position so that they do not generate any lift, they will still produce pressure and friction drag, which is the profile drag that the flight muscles must overcome to rotate the wings for flapping. In flapping flight, the magnitude of the profile drag depends on both the rotation speed of the wings, which changes along the wingspan from zero at the wing root to maximum at the wing tip, and the forward speed. There is a tradeoff between the profile drag coefficient (decreasing with increasing forward speed) and the profile drag owing to forward speed, resulting in a near-constant profile power in the range of typical cruising speeds. Profile power ( $P_{pro}$ ) is provisionally calculated as

$$P_{pro} = XP_{am},$$

where  $X$  is a constant and  $P_{am}$  is the minimum of the sum of induced and parasite power. At fast speeds, the profile power will increase with speed in a similar way as the parasite power and so this approximation cannot be extrapolated outside the range in which it is valid.

Mechanical power of bird flight The total mechanical power ( $P_{mech}$ ) required to fly is

$$P_{mech} = P_{ind} + P_{par} + P_{pro}$$

which is a function of forward speed, bird morphology and air density.

## ACKNOWLEDGMENTS

We thank two reviewers for their constructive comments on the manuscript. This work is partially supported by an NSF grant (0540745) to RW.

## REFERENCES

1. Rayner JMV. Estimating power curves of flying vertebrates. *J Exp Biol* 1999; 202:449-461.
2. Poore SO, Sanchez-Haiman A, Goslow GE. Wing upstroke and the evolution of flapping flight. *Nature* 1997; 387:799-802.
3. Bryant DM. Energy expenditure in wild birds. *Proc Nutr Soc* 1997; 56:1025-1039.
4. Ward WS, Möller U, Rayner JMV, Jackson DM, Bilo D, Nachtigall W, Speakman JR. Metabolic power, mechanical power and efficiency during wind tunnel flight by the European starling *Sturnus vulgaris*. *J Exp Biol* 2001; 204:3311-3322.
5. Dial KP, Biewener AA, Tobalske BW, Warrick DR. Mechanical power output of bird flight. *Nature* 1997; 390:67-70.
6. Alexander RM. The U, J and L of bird flight. *Nature* 1997; 390:13-13.
7. Tobalske BW, Hedrick TL, Dial KP, Biewener AA. Comparative power curves in bird flight. *Nature* 2003; 421:363-366.
8. Hedenström A. Aerodynamics, evolution and ecology of avian flight. *Trends Eco Evo* 2002; 17:415-422.
9. Kvist A, Klaasen M, Lindström A. Energy expenditure in relation to flight speed: what is the power of mass loss rate estimates? *J Avian Biol* 1998; 29:485-498.
10. Pennycuik CJ. Bird Flight Performance: A Practical Calculation Manual. Oxford University Press (1989).
11. Ellington CP. Limitations on animal flight performance. *J Exp Biol* 1991; 160:71-91.
12. Ma CX, Casella G, Wu RL. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. *Genetics* 2002; 161:1751-1762.
13. Lou X-Y, Casella G, Littell RC, Yang MCK, Johnson JA, Wu RL. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. *Genetics* 2003; 163:1533-1548.
14. Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA (1998).
15. Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002; 18:19-24.
16. Hästbacka J, de la Chapelle A, Mahtani M, Clines G, Reeve-Daly MP, Daly M *et al*. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 1994; 78:1073-1087.
17. Wang ZH, Wu RL. A statistical model for high-resolution mapping of quantitative trait loci determining HIV dynamics. *Stat Med* 2004; 23:3033-3051.
18. Wu RL, Ma CX, Chang M, Littell RC, Wu SS, Huang M, Wang M, Casella G. A logistic mixture model for detecting major genes governing growth trajectories. *Genet Res* 2002; 79:235-245.
19. Kirkpatrick M, Lofsvold D, Bulmer M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 1999; 124:979-993.
20. Pletcher SD, Geyer CJ. The genetic analysis of age-dependent traits: Modeling the character process. *Genetics* 1999; 153:825-835.
21. Jaffrezix F, Pletcher SD. Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics* 2002; 156:913-922.
22. Sillanpää MJ, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* 1999; 151:1605-1619.
23. Zhao W, Wu RL, Ma C-X, Casella G. A fast algorithm for functional mapping of complex traits. *Genetics* 2004; 167:2133-2137.
24. Patil N, Berno AJ, Hinds DA, Barrett WA *et al*. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294:1719-1723.
25. Dawson E, Abecasis GR, Bumpstead S, Chen Y *et al*. A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 2002; 418:544-548.
26. Gabriel SB, Schaffner SF, Nguyen H, Moore JM *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; 296:2225-2229.