Comparison of techniques for correlating survival and gene expression data from wild salmon

Edd Hammill1,2, Janelle M.R. Curtis2, David A. Patterson3, Anthony P. Farrell4, Thomas Sierocinski4, Paul Pavlidis5, Scott G. Hinch6, Kristi Miller2

1Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada
2Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada
3Fisheries and Oceans Canada, Cooperative Resource Management Institute, School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC, Canada
4Department of Zoology and Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC, Canada
5Department of Psychiatry, Centre for High-throughput Biology, University of British Columbia, Vancouver, BC, Canada
6Centre for Applied Conservation Research, Department of Forest Sciences, University of British Columbia, Vancouver, BC, Canada

Accepted for publication September 13, 2011

Abstract – In laboratory and field studies of survival, one of two alternative analytical techniques is often used to estimate survival rates and identify covariates, namely parametric survival analysis or Cormack–Jolly–Seber models. These techniques differ in algorithms and assumptions of the data. They also tend to be used under different circumstances depending on whether the intention is to demonstrate group-specific differences or to predict survival variables. Here, we apply and compare both analytical techniques in a study that couples functional genomics with biotelemetry to ascertain the role of physiological condition on survival of adult sockeye salmon (Oncorhynchus nerka) migrating in the Fraser River, British Columbia, which builds on the growing concern over the decline in numbers of spawning fish. Herein, we show a high level of quantitative and qualitative agreement between the two analytical methods, with both showing a strong relationship exists between the genomic signature that accounts for the largest source of variance in gene expression among individuals and survival in one of the three populations assessed. This high level of agreement suggests the data and the approaches are generating reliable results. The novel approach used in our study to identify physiological processes associated with reduced fitness in wild populations should be of broad interest to conservation biologists and resource managers as it may help reduce the uncertainty associated with predicting population sizes.

Key words: sockeye; salmon; survival analysis; genomics

Introduction

Quantifying the survival patterns of wild individuals is a key to understanding the dynamics of populations (Maynard-Smith 1974). Knowledge of natural survival rates is especially important for any species targeted by commercial harvesting as it enhances the ability to set sustainable catch limits (Walters & Martell 2004), and is essential for meeting conservation and economic objectives. As small changes in survival rate can have disproportionately large effects on population dynamics (Caswell 2000), survival rates need to be predicted within narrow margins of error. In the current study, we compare and contrast two alternative methods for analysing survival data and evaluate the robustness of physiological predictors of survival in wild sockeye salmon (Oncorhynchus nerka).

Individual health and survival are determined through a complex interplay between genetics and environmental conditions (Kingsolver 1995; Siegel 1995; Lohmus et al. 2010). As environmental variables are reasonably easy to measure, they are often explored as potential correlates of survival in wild populations (Sedinger et al. 1995; Walters & Reich 1996; Beckage & Clark 2003; Pörtners & Farrell 2008). For example, studies of chinook and coho salmon...
correlate a wide range of oceanographic variables with survival (Beamish et al. 2010). Nevertheless, retrospective analyses suggest that environmental covariates are poor predictors of survival in salmon population-recruitment models (e.g., sea surface temperature, Haeseker et al. 2008). Should physiology be related to performance it will act as the key link between the environment and an individual’s survival (Pyper et al. 2005), direct measurement of salmon physiology and condition may provide a more effective predictor of survival than indirect measures of environmental conditions.

Physiological traits including cortisol, plasma lactate and osmoregulatory function have been individually related to rates of wild salmon mortality (Cooke et al. 2006; Crossin et al. 2008; Miller et al. 2011), yet, to date, none have been incorporated into salmon population-recruitment models as predictive covariates of recruitment or survival. Although these studies provide insight into some of the mechanisms influencing salmon health, the measurement of individual physiological traits provides a limited indication of the overall condition of an individual. In contrast, the rapidly developing field of functional genomics, however, provides opportunities to simultaneously measure a wide range of physiological conditions and produce an estimation of the overall health of an individual. The use of microarrays allows the investigation of patterns of gene expression in tens of thousands of genes (Salem et al. 2006). Previous research has shown that genomic signatures are correlated with premature in-river mortality of migrating adult sockeye salmon (Miller et al. 2011). Because of the importance of this discovery, and its potential economic significance, the current study uses genomics and biotelemetry data to compare the results of two alternative methods to analyse and relate survival data to patterns in gene expression.

Direct measurements of survival rates from wild populations are hard to obtain as they require either repeated encounters with marked individuals or some form of continuous monitoring (Gaillard et al. 1993; Bjornsdal et al. 2003; Rowat et al. 2009). Acquisition of reliable data is especially hard in species that migrate thousands of kilometres through diverse environments (Bjornsdal et al. 2003), which is the case for anadromous salmon. In recent years, improvements in tag technology and widely dispersed receiver arrays have made the reliable tracking of large aquatic vertebrates possible (Welch et al. 2009).

Once collected, the analysis of survival data is more complex than many other statistical procedures because the response variable is made up of two components, whether or not an event occurred (i.e., whether an individual survived or died) and time to the event or end of study. In addition, as the variance of the response variable increases during the course of the experiment (Harrell 2001), the assumption of constant variance required for regular regression analysis is violated. Circumventing these difficulties, two methods of analysing survival data have been widely used for laboratory and field experiments. Parametric survival analysis (PSA) (an extension of generalised linear models) is designed specifically to look at “time to event” data (Harrell 2001; Crawley 2007) and is often used in laboratory studies where precise data on time of death is easily acquired (Hammill & Beckerman 2010). Analysis of survival data can also be achieved by treating them as mark–recapture observations (Loison et al. 1999). One of the most popular methods of analysing mark–recapture data utilises the Cormack–Jolly–Seber (CJS) model. This model estimates two key parameters, apparent survival (Φ), an estimate of the number of individuals surviving to a certain time or location, and the probability that a surviving individual is detected (P) (reviewed in Lebreton et al., 1992). The CJS model is often used in field studies where knowing whether or not an individual was alive at a particular time is of greater concern than the exact time of death, which is more difficult to determine in situ.

In recent years, poor in-river survival of sockeye salmon in the Fraser River, British Columbia (BC), has become a serious concern with unusually high variance in mortality rates observed as adults migrate back towards spawning grounds (Cooke et al. 2004; Farrell et al. 2008). Sockeye salmon are the most economically and culturally important salmonid species in the Fraser River watershed (Jacoby et al. 2010), and recent declines have prompted a Canadian judicial enquiry into the causes (Cohen Commission 2009). Understanding these unprecedented mortality and recruitment events and enhancing the predictability of en-route migration and prespawning mortality are high research priorities for managers tasked with ensuring long-term sustainability. The goal of the present study was to compare and contrast two methods of analysing adult survival data from wild sockeye salmon and relating survival to patterns in gene expression (i.e., physiological condition). Should the two analytical methods qualitatively and quantitatively agree, we can have more confidence in the potential to predict adult sockeye salmon survival on the basis of their physiology during migration back to spawning grounds.

Materials and methods

Our study contrasted two methods of assessing survivorship using physiological information and is based on information on physiological condition, movement, and fate obtained by coupling biotelemetry, biopsy sampling and genomics. As most of the
field and genomics methods have been published elsewhere (Martins et al. 2011; Miller et al. 2011), we only briefly describe the radio-tracking and microarray methods used to generate those datasets that formed the basis of our statistical comparison.

Study area, capturing and radio-tracking

Sockeye salmon migrating up the Fraser River return to their natal tributaries to spawn, and fish from each of these tributaries represent different subpopulations or population complexes (Groot & Margolis 1991). We monitored fish from three population complexes: Scotch Creek \(n = 21\), Chilko \(n = 24\) and Adams \(n = 32\). Capture, tagging, biopsy and telemetry data acquisition were carried out as detailed in Martins et al. (2011) with fish tagged at Whonnock, BC, as they entered the Fraser River in July and August of 2006. Survival and migration patterns were monitored between Mission, BC, and the location of each fish’s natal spawning ground (deduced from genetic population identification) between 14 July and 31 August 2006. Any fish reported as a fishery capture was removed from the analysis because its mortality was not natural and likely unrelated to pre-existing physiological condition (Cooke et al. 2009). Salmon migration was detected at 16–24 monitoring stations (depending on route) placed at strategic sites along the Fraser River (Fig. 1). At each site, a receiver (models SRX400, SRX400A or SRX600; Lotek Wireless, BioSonics Telemetry LP, 4027 Leary Way NW, Seattle, WA 98107, USA) with two or three antennae (3- or 4-element Yagi; Max, Inc., Hannover Park, IL, USA or Grant Systems Engineering, Inc. of King City, ON, Canada) was deployed >10 m above water level.

Tissue collection for microarray and population determination

A small piece (<4 mm) of gill filament tip was taken for microarray analysis as telemetry tags were implanted. A clip of adipose fin was also removed for genetic population identification (Beacham et al. 2006). Fish used in our study were assigned to one of three population complexes: Scotch Creek (Early Shuswap complex), Adams (included some genetically similar Late Shuswap complex populations), and Chilko (a Summer run population). Elevated mortality was not associated with gill biopsy sampling for tagged fish (Cooke et al. 2005; Robichaud & English 2007).

A 16,006 feature microarray was used to assess the physiological condition of salmon at the time of Fraser River entry. Features with poor quality control or no signal (fold change across samples <2 times background) were flagged; spots with ≥50% flags across individual arrays were removed from the remainder of the analysis. Following the removal of these spots, we were left with expression data on 9236 genes. Isolation and purification of nucleic acids, as well as collection of microarray data, were carried out as in the study by Miller et al. (2009). Normalisation procedures used are described in the study by Miller et al. (2011).

Analysis of gene expression data

We used gene expression data to assign sampled fish to categories that putatively reflect differences in physiological condition (i.e., genomic signature groups). To

![Fig. 1. Location of the detection stations within the Fraser River watershed.](image)
achieve this, we first ran a principal component analysis on the array data of all 9236 genes to discover unique physiological signatures in the data. We then identified the subsets of genes with high (>50% of maximum) absolute rotation values (the gene’s effect on the principal component) for the first three principal components and independently used these to cluster fish into genomic groups. This use of a subset of genes was employed to improve the signal/noise ratio. Below 50%, the information contributed per gene drops substantially. As principal component 3 explains considerably less of the overall variation than principal component 1, the genes involved in grouping the fish based on the third component are having a weaker overall effect than those grouping on principal component 1. Hierarchical clustering was employed to group fish according to the subset of genes associated with principal components 1–3; percentage support for each split was determined by bootstrapping the data 10,000 times (Suzuki & Shimodaira 2006). We then looked for differences in survival between the genomic groups using two different analytical approaches: PSA and Cormack–Jolly–Seber modelling.

Parametric survival analysis

In PSA, the response variable consisted of whether or not an individual reached the spawning ground, and if not, how long it survived. Fish that passed the last receiver station adjacent to their spawning ground were assigned a value corresponding to the day the last survivor arrived at the last detection station. The day the last survivor arrived was also treated as the last day of the study, specific to each population. Fish that did not reach the last detection station never survived to the final day of the study, and their time of death was the last day they were detected. We then looked for differences in this response variable between the genomic groups. The survival analysis models were produced in R (Team 2011) and utilised parametric survival models with a time-specific hazard function and a lognormal error distribution. We opted for this error distribution following comparisons of residual plots, choosing the structure that produced the minimal error deviance. Because the natal population (Scotch, Chilko, Adams) and sex of each fish were known, these factors were included in the analysis and tested for significant interaction terms. As survival on the spawning grounds was not monitored, our data were restricted to assessments of en-route survival.

Mark–recapture models

Program MARK (White & Burnham 1999) to fit a Cormack–Jolly–Seber model (CJS model) was used to estimate two types of parameters: apparent survival, $\Phi$, and detection probability, $P$. The first detection station encountered by all fish was at Mission (Fig. 1); when a fish was detected at this station, it was deemed ‘marked’ and used in the study. We then selected a further three stations (the final one being the station closest to the natal spawning ground), restricting our analysis to a subset of evenly distributed stations to reduce the number of parameters that needed to be estimated. Every time, a fish was detected at one of these stations it was recorded as ‘recaptured’. The distance between stations was also used in the CJS model so that it could be accounted for when estimating $\Phi$. For the Scotch and Adams populations, the detection stations after Mission were located at (length of river section in parenthesis) (Fig. 1) Hell’s Gate (134 km), Spence’s Bridge (94 km) and Little River (105 km). For Chilko fish, selected tracking stations were at Hell’s Gate (134 km), the Seton Confluence (113 km) and the Chilcotin Confluence (153 km). A total of 15 mark–recapture models were fitted to each population. In the full model, denoted $\Phi(lgs), P(l), \Phi (.) , P (.)$. The other 13 models were composed of reduced versions of the full model. We used Akaike’s corrected information criterion (AICc) (Burnham & Anderson 2002) to find the most parsimonious model to describe the data for each population. The full model is always the best fit to the data in terms of deviance; however, many parameters may account for very little variation (Burnham & Anderson 2002). AICc takes into account both the fit of a model and the number of parameters it contains to find the model that best balances parsimony and explanatory power. In addition, we calculated Akaike weights (AICw) to find the relative support for each model (Wagenmakers & Farrell 2004) and then used model averaging (average parameter estimates across models weighted by AICw) to obtain estimates of apparent survival and detection probabilities.

Results

Genomic signals

Principal component 1 explained 19.2% of the total variation in the genomic data. When the genes were ranked in order of their absolute rotation value for PC1, 9% (837) of the 9236 genes included in the
Comparison of techniques for correlating survival and gene expression data

Table 1. Results of the mark–recapture analysis, showing AICc and AICw for the top 10 models for each population. \( \Phi \) refers to survival, \( P \) refers to detection probability.

<table>
<thead>
<tr>
<th>Model</th>
<th>( \Delta \text{AIC} )</th>
<th>AICc</th>
<th>Chilko</th>
<th>Model</th>
<th>( \Delta \text{AIC} )</th>
<th>AICc</th>
<th>Adams</th>
<th>Model</th>
<th>( \Delta \text{AIC} )</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 0.00 )</td>
<td>( 0.00 )</td>
<td>0.35</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 0.00 )</td>
<td>( 0.00 )</td>
<td>( 0.02 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 2.89 )</td>
<td>( 1.97 )</td>
<td>( 1.97 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 2.11 )</td>
<td>( 0.12 )</td>
<td>( 0.12 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 3.33 )</td>
<td>( 2.20 )</td>
<td>( 2.20 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 2.11 )</td>
<td>( 0.12 )</td>
<td>( 0.12 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 3.44 )</td>
<td>( 5.94 )</td>
<td>( 5.94 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 2.35 )</td>
<td>( 0.11 )</td>
<td>( 0.11 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 5.03 )</td>
<td>( 4.11 )</td>
<td>( 4.11 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 4.55 )</td>
<td>( 0.04 )</td>
<td>( 0.04 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 7.28 )</td>
<td>( 4.61 )</td>
<td>( 4.61 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 4.61 )</td>
<td>( 0.04 )</td>
<td>( 0.04 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 8.65 )</td>
<td>( 5.01 )</td>
<td>( 5.01 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 5.01 )</td>
<td>( 0.03 )</td>
<td>( 0.03 )</td>
<td>( \phi (g</td>
</tr>
<tr>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 10.59 )</td>
<td>( 6.41 )</td>
<td>( 6.41 )</td>
<td>( \Phi (g</td>
<td>P_l) )</td>
<td>( 6.41 )</td>
<td>( 0.01 )</td>
<td>( 0.01 )</td>
<td>( \phi (g</td>
</tr>
</tbody>
</table>

AICc, Akaike’s corrected information criterion; AICw Akaike’s corrected information weights. The letters in parenthesis refer to factors by which the parameters are allowed to vary: \( g \) = genomic group, \( s \) = sex, \( l \) = location (i.e., section of the river), \( . \) = parameter is constant. For each population the first model in which survival varies with, or is involved in an interaction with genomic group is highlighted.

### Results

#### Principal Component Analysis (PCA)

The principal component analysis (PCA) had an absolute rotation value >50% of the maximum. Clustering these 837 genes resulted in a first split that divided fish into two clusters of reasonably equal size (Fig. 2; group 1, \( n = 33 \), group 2, \( n = 44 \)) with 92% support (approximately unbiased value). Principal component 2 accounted for 8.4% of the variation and principal component 3 accounted for 4.2%. Neither of these two components produced clear groups through bootstrapped hierarchical clustering (<50% support for first splits).

#### Parametric Survival Analysis

Using the groups generated from PC1, a significant interaction was found between genomic group and population \( (F_{2,70} = 8.70, P = 0.012) \), meaning the relationship between genomic group and survival differed among the salmon population complexes (Fig. 3). The main effects of population and group were also included in the model, but neither effect was significant (both \( P < 0.1 \)). No significant differences in survival were observed between sexes \( (F_{1,71} = 0.067, P = 0.8) \). Because of the significant group by population interaction term, we carried out the survival analysis on the three populations individually while also employing a Bonferroni correction, reducing the corrected \( \alpha \) at which the null hypothesis could be rejected to 0.0167. For the Scotch population complex, individuals that clustered into genomic group 2 survived for a significantly shorter length of time \( (F_{1,19} = 8.77, P = 0.005) \). There was no significant difference between fish clustered into the two genomic groups for the Chilko population \( (F_{1,22} = 3.34, P = 0.07) \) or the Adams population \( (F_{1,30} = 0.008, P = 0.93) \).

There were no significant differences in survival between the groups generated from PC2 and PC3. There were no significant interactions between group and sex or population, and the 3-way interaction among all three factors was also not significant (all \( P > 0.05 ) \).
Comparison of methods for estimating survival rates in PC1

Scotch
The two statistical techniques produced quantitatively similar estimates of survival. The mean number of days a surviving fish took to reach the spawning grounds from Mission was 17. From the PSA, on this day the proportion of fish surviving was 0.60 (95% CI = 0.36–0.78) for genomic group 1 and 0.12 (95% CI = 0.02–0.31) for genomic group 2. The CJS model produced an estimate of 0.61 for genomic group 1 (95% CI = 0.37–0.83) and 0.06 (95% CI = 0.00–0.48) for group 2. For both genomic groups, the mean produced using survival analysis is within the confidence limits generated from the CJS model and vice versa (Fig. 4).

Chilko
The two statistical tests again produced quantitatively similar estimates of survival. Fish that survived to the final receiver took an average of 14 days to arrive. At this date, the proportion surviving (according to PSA) was 0.28 (95% CI = 0.12–0.49) for genomic group 1 and 0.41 (95% CI = 0.2–0.66) for genomic group 2. CJS generated an estimate of survival equal to 0.18 (95% CI = 0–0.45) for group 1 and 0.21 (95% CI = 0–0.39) for group 2. As both the confidence limits obtained from the different analyses contain the estimate from the other analysis (Fig. 4), they are quantitatively similar.

Adams
A surviving fish took on average 24 days to reach the spawning ground. At day 24, the estimate of proportion surviving according to PSA was 0.24 (95% CI = 0.17–0.47) for genomic group 1 and 0.22 (95% CI = 0.16–0.43) for group 2. Our CJS estimate was 0.32 (95% CI = 0.13–0.50) for group 1 and 0.28 (95% CI = 0.11–0.41) for group 2. Therefore, we can say the two estimates quantitatively agree (Fig. 4).

Discussion
By combining biotelemetry data and functional genomics, we provide evidence that different patterns...
of gene expression in sockeye salmon are significantly related to survival in at least one important population complex (Scotch). The high level of quantitative and qualitative agreement between the PSA and mark–recapture modelling provides evidence that our telemetry and genomics data are sufficient to reveal the effect of physiological condition on patterns of survival in wild migrating salmon. The results also highlight the importance of accounting for population-specific differences when predicting en-route survival and incorporating such information into management decisions (Macdonald et al. 2010).

The high level of agreement between the two analytical techniques comes in spite of fundamental differences in the algorithms they employ. Goodness-of-fit testing of the CJS models showed reasonable fits to the data (Lebreton et al. 1992), and the most parsimonious models estimated a capture probability not significantly different from 1. Of the two methods, the CJS models make fewer assumptions about the
error structure of the data by simply looking at whether or not an individual survived to a particular detection station. Hence, the CJS model may be generally more robust than PSA. The former analytical technique could therefore be a better method to assess differences between populations (Burnham & Anderson 2002; Crawley 2007). However, PSA is able to give a more precise estimate of survival time or the chances of surviving to a particular day (Crawley 2007; Hammill & Beckerman 2010) and produces modelled data that make it possible to graphically represent survival differences over the duration of the experiment. The most appropriate method depends on whether the intention is increased robustness to statistical assumptions or precision, but either appears suitable for demonstrating the effect of physiological condition on survival.

The genomic signature associated with PC1 explained the largest single source of variation in the data, and all three populations were represented on both ends of the PC1 distribution. However, the signature did not have the same effect on survival in all three populations. The PC1 genomic signature clearly adversely affected Scotch fish. A functional analysis of the microarray data conducted in a previous study has suggested that the physiological signal associated with higher mortality is related to a viral infection (Miller et al. 2011).

Considering the close geographic proximity of the Adams and Scotch Creek spawning grounds, we may initially expect the two populations to behave in a similar manner. However, fish from the Adams and Scotch Creek populations are temporally isolated, tending to arrive in the river at different times (Groot & Margolis 1991). The Scotch Creek population is an ‘early summer’ run, with fish tending to return in mid-July. The Adams population is a ‘late’ run, with the majority of the fish returning after mid-September. This would mean that under normal conditions, the two populations would be exposed to very different environmental conditions, with Scotch Creek fish being exposed to much warmer temperatures. The fish used in this study were all tagged and monitored at the same time, between 14 July and 31 August, and fish from the three populations were distributed within this time window. As a result, the fish from the Adams populations were migrating early. These ‘early entry’ migrants tend to be compromised before they enter the rivers, through pathogenic infections such as Parvicipausa and chronic thermal tolerance issues (Wagner et al. 2005; Hinch 2009). This would mean we would expect the Adams fish found in this study to be in poor condition and have low levels of survival as they would be entering in a compromised condition, and then be exposed to temperatures approximately 5 °C greater than their ideal thermal range. Evidence for the poor condition of the Adams fish can be seen in Fig. 4, and survival rates for Adams fish from both genomic groups were low, similar to those of group 2 Scotch Creek fish. We therefore suspect that the Adams fish from both groups were unhealthy, and this was the reason we did not observe a significant difference in survival between them. We therefore may need to interpret the results for the Scotch Creek and Adams populations slightly differently. For the Scotch Creek population, we sampled the population during their normal migration time, and were therefore able to isolate a group containing a signal of poor health, and found a relationship between this signal and survival. For the Adams population, we may have only sample ‘early’ unhealthy fish. Although we were able again to isolate the same genomic signal as we found in Scotch population, all the Adams fish were in poor condition and so all had poor survival. If we had tagged fish between mid- and late September, we would potentially have sampled ‘normal’ Adams fish and then observed a relationship between genomic group and survival.

The same genomic signature has been identified in sockeye salmon sampled in the marine environment >200 km before fish reached the Fraser River (Miller et al. 2011). In ocean-tagged fish tracked from Hells Fig. 4. Estimates and confidence limits of rates of survival to spawning ground estimated using the 2 different analytical techniques. Open circles (○) are estimates obtained using parametric survival analysis, filled circles (●) are estimates obtained from the best Cormack–Jolly–Seber model.
Gate to spawning grounds, the most stressful region of river migration, this signature was associated with reduced survival of Adams fish. Hence, effects on survival for some populations may depend on the length of time they have been stressed (ocean tagging followed survival of fish long after they showed signs of the signature, tracking them for a total of 1–2 months) and/or the presence of other stressors in the river such as temperature. As the fish used in this study were not sampled until they reached freshwater, many individuals carrying the signature associated with poor survival may have died before being sampled. Moreover, this same genomic signature was associated with prespawning mortality at the Weaver Creek spawning ground (Hruska et al. 2010), suggesting that premature losses associated with this signature may continue once fish reach spawning areas. This may be an important factor in the lack of a significant result for the Chilko population as they may simply have not been tracked long enough to uncover an effect. Chilko fish were tracked for the shortest distance in the river, and the last receiver was still 1-week travel time from their spawning ground, whereas the last receiver was only one-day travel time from Adams and Scotch spawning grounds. Hence, we suspect that our sampling likely underestimated total prespawning losses and to a greater extent for Chilko fish.

Population-specific differences in correlations between physiological variables and survival have been previously demonstrated in Fraser River sockeye salmon, with biomarkers associated with energy, osmoregulation, reproductive maturation and stress all associated with survival of certain populations in particular years (Cooke et al. 2006). Between-population results however have shown a lack of consistency, with high levels of plasma cortisol being correlated with failure to reach the spawning grounds in Quesnel and Chilko populations, while low levels have been associated with failure to reach the river in Nechako fish (Cooke et al. 2006). Interestingly, the genomic signature associated with premature mortality herein also contains elements suggestive of osmoregulatory dysfunction (Miller et al. 2011) and is associated, at least in Scotch Creek fish, with highly elevated mortality in the lower river, consistent with an earlier study (Cooke et al. 2006).

Management implications
The relationship between salmon condition and survival could have important management implications, but there is clearly more work to be performed to fully explore relationships in different years, populations and at different phases of migration. As the relationship between different genomic signals and survival becomes better understood, functional genomics may be a suitable starting point from which to develop predictive biomarkers that could be used to assess the condition of large numbers of returning salmon. These data could then be incorporated into existing assessment and decision-support models (e.g., Pacific Salmon Commission in-season management adjustment models).

Future directions and broader implications
We have shown that the two commonly used methods of analysing survival data provide strikingly similar results when applied to our genomics-telemetry data, suggesting that both the data and the methods are robust. The use of telemetry has some clear advantages over standard mark–recapture techniques, which can only be analysed using the CJS method. Although expensive, the increased precision of modern telemetry techniques that have high detection probabilities and detect individuals at multiple locations substantially improves the precision of the data acquired. This improved precision means more reliable estimates of ‘day of death’ can be obtained. We believe this improved precision has aided our ability to identify physiological factors that negatively impact salmon survival in the wild. This approach is transportable to a broad range of taxa, and we hope it will become a benchmark for future work elucidating the role of physiological condition on fitness of wild populations.

Acknowledgements
This research was funded through a Genome BC AGIP project awarded to K. Miller, J. Curtis, S. Hinch, A. Farrell and P. Pavlidis. Funding was also provided by an NSERC Strategic Grant to S. Hinch and A. Farrell, and grants from the Southern Endowment Fund of the Pacific Salmon. EH was supported by a NSERC Visiting fellowship in a government laboratory award. All research was approved by the University of British Columbia animal care committee. M. Donaldson, I. Olsson, G. Crossin, K. Hanson, R. Alexander, D. Robichaud, S. Tyerman, D. Welch and J. Hills conducted tagging biopsies. This MS was improved through discussion with Eduardo. Martins and Brian Ma. We would like to thank all in the Pacific Biological Station molecular genetics group.

References
Beamish, R.J., Sweeting, R.M., Lange, K.L., Noakes, D.J., Preikshot, D. & Neville, C.M. 2010. Early marine survival of coho salmon in the Strait of Georgia declines to very low
Hammill et al.


Comparison of techniques for correlating survival and gene expression data


