Appendix A. Proof of Results in the Main Text

In this appendix we provide proofs for the theoretical results in the main text.

**Proof** [Proposition 1] Given $S, S' \in \mathcal{S}_n$, we obtain from the pseudo-metric properties of $d_{\mathcal{P}}$ each of (i) positivity $d_{\mathcal{S}}(S, S') = d_{\mathcal{P}}(\kappa(S), \kappa(S')) \geq 0$, (ii) pseudo-identity $d_{\mathcal{S}}(S, S) = d_{\mathcal{P}}(\kappa(S), \kappa(S)) = 0$, (iii) symmetry $d_{\mathcal{S}}(S, S') = d_{\mathcal{P}}(\kappa(S), \kappa(S')) = d_{\mathcal{P}}(\kappa(S'), \kappa(S)) = d_{\mathcal{S}}(S', S)$ and (iv) the triangle inequality $d_{\mathcal{S}}(S, S'') = d_{\mathcal{P}}(\kappa(S), \kappa(S'')) \leq d_{\mathcal{P}}(\kappa(S), \kappa(S')) + d_{\mathcal{P}}(\kappa(S'), \kappa(S'')) = d_{\mathcal{S}}(S, S') + d_{\mathcal{S}}(S', S'')$, for all $S, S', S'' \in \mathcal{S}_n$. 
Suppose now that $\kappa$ is injective and $d_P$ is a metric on $\mathcal{P}$. Then if it holds that $d_S(S, S') = d_P(\kappa(S), \kappa(S')) = 0$, it follows that $\kappa(S) = \kappa(S')$ which (from assumption on $\kappa$) implies $S = S'$ in $S_n$. Thus under these additional assumptions, $d_S$ is a metric on $S_n$. $lacksquare$

**Proof** [Proposition 2] The non-negativity, symmetry and sub-additivity properties of $d_P$ are clear, so all that remains is to establish that $d_P(\pi, \pi') = 0$ implies $\pi = \pi'$. From the definition of $\mathcal{P}$, both $\pi$ and $\pi'$ are continuous on $Z_2$. The result is then immediate from the fact that, since $\pi$ and $\pi'$ are continuous and $Z_2$ is compact, then $\int_{Z_2} |\pi(z') - \pi'(z')|^2 \, d\Lambda_2(z') = 0$ implies $\pi$ and $\pi'$ must be identical as functions on $Z_2$. $lacksquare$

**Proof** [Proposition 3] Observe that, using Prop. 2 for sub-additivity of the metric $d_P$,

\[
d_S(S^{(n)}, \tilde{S}^{(n)}) = d_P(\kappa(S^{(n)}), \kappa(\tilde{S}^{(n)})) \\
\leq d_P(\kappa(S^{(n)}), \pi) + d_P(\pi, \tilde{\pi}) + d_P(\tilde{\pi}, \kappa(\tilde{S}^{(n)})) \\
= d_P(\pi, \tilde{\pi}) + \|\pi - \kappa(S^{(n)})\|_{L^2(\Lambda_2)} + \|\tilde{\pi} - \kappa(\tilde{S}^{(n)})\|_{L^2(\Lambda_2)}
\]

Since $\kappa$ is consistent we have $\|\pi - \kappa(S^{(n)})\|_2 = o_P(1)$ and $\|\tilde{\pi} - \kappa(\tilde{S}^{(n)})\|_2 = o_P(1)$. This completes the proof. $lacksquare$

**Proof** [Proposition 4] This proof extends the simpler proof given for the univariate case in Theorem 6.11 of Wassermann (2006). For convenience, and without loss of generality, we suppose that $Z_2 = [0, 1]^2$. It will be convenient in this section to re-assign the notation $z$ as a dummy variable in $Z_2$ (instead of in $Z_P$). Let

\[
p_{i,j} = \int_{B_{i,j}} \pi d\Lambda_2
\]

be the probability mass assigned to

\[
B_{i,j} = \left[ z_{\min} + \left( z_{\max} - z_{\min} \right) \frac{i-1}{M}, z_{\min} + \left( z_{\max} - z_{\min} \right) \frac{i}{M} \right] \\
\times \left[ z_{\min} + \left( z_{\max} - z_{\min} \right) \frac{j-1}{M}, z_{\min} + \left( z_{\max} - z_{\min} \right) \frac{j}{M} \right]
\]

so that, from binomial properties, the mean and variance of the histogram estimator $\kappa(S^{(n)})(z)$ at the point $z \in Z_2$ are

\[
m(z) = \frac{p_{i,j}}{h^2} \\
v(z) = \frac{p_{i,j}(1 - p_{i,j})}{nh^4}
\]

Let $b(z) = m(z) - \pi(z)$ denote the bias of the histogram estimator. The mean square of the error $\pi(z) - \kappa(S^{(n)})(z)$ at a point $z' \in Z_2$ can be bias-variance decomposed:

\[
\mathbb{E}\{[\pi(z) - \kappa(S^{(n)})(z)]^2\} = b(z)^2 + v(z)
\]
The aim is to obtain independent bounds on both the bias and variance terms next.

To bound the bias term, Taylor’s theorem gives that, for \( z, z' \in B_{i,j} \),

\[
\pi(z') = \pi(z) + (z' - z) \cdot \nabla \pi(z) + \frac{1}{2} (z' - z) \cdot R_{i,j}(z) (z' - z)
\]  

(7)

where the remainder term satisfies

\[
\|R_{i,j}(z)\|_{\text{max}} \leq \sup_{z'' \in B_{i,j}} \|\nabla \nabla^\top \pi(z'')\|_{\text{max}} \quad \text{(Taylor)}
\]

\[
\leq \sup_{z'' \in Z_2} \|\nabla \nabla^\top \pi(z'')\|_{\text{max}} < \infty \quad \text{(continuous on compact domain)}.
\]

Here \( \|M\|_{\text{max}} = \max\{M_{i,j}\} \) and \( \nabla \nabla^\top \pi \) denotes the Hessian, which exists since \( \pi \) is twice continuously differentiable in \( Z_2 \). Thus for \( z \in B_{i,j} \), integrating (7):

\[
\int_{B_{i,j}} \pi(z')d\Lambda_2(z') = h^2 \pi(z) + h^2 \left( \frac{h}{2} \left[ \frac{2i - 1}{2j - 1} \right] - z \right) \cdot \nabla \pi(z) + E_{i,j}(z)
\]

where the new remainder term can be bounded:

\[
|E_{i,j}(z)| = \left| \frac{1}{2} \int_{B_{i,j}} (z' - z) \cdot R_{i,j}(z) (z' - z) d\Lambda_2(z') \right|
\]

\[
\leq \frac{1}{2} \int_{B_{i,j}} \|z' - z\|^2 d\Lambda_2(z') \times \sup_{z'' \in Z_2} \|\nabla \nabla^\top \pi(z'')\|_{\text{max}}
\]

\[
\leq 8h^4 \sup_{z'' \in Z_2} \|\nabla \nabla^\top \pi(z'')\|_{\text{max}} = C h^4
\]

where the constant \( C \) is independent of \( z \) and \( i, j \). The number 8 (which is not sharp) is obtained from trivial but tedious computation of the integral in (8) and bounding each term in the result. Now, for \( z \in B_{i,j} \), the bias is expressed using (7) as

\[
b(z) = \frac{1}{h^2} \int_{B_{i,j}} \pi(z')d\Lambda_2(z') - \pi(z)
\]

\[
= \left( \frac{h}{2} \left[ \frac{2i - 1}{2j - 1} \right] - z \right) \cdot \nabla \pi(z) + \frac{1}{h^2} E_{i,j}(z).
\]
Now we integrate this expression over $x \in B_{i,j}$:

\[
\int_{B_{i,j}} b^2 d\Lambda_2 = \int_{B_{i,j}} \left\{ \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) + \frac{1}{h^2} E_{i,j} \left( z_2 \right) \right\}^2 d\Lambda_2(z)
\]

\[
\leq \int_{B_{i,j}} \left\{ \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right\}^2 d\Lambda_2(z)
\]

\[
+ 2 \int_{B_{i,j}} \left| \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right| \frac{1}{h^2} |E_{i,j}(z)| d\Lambda_2(z)
\]

\[
+ \int_{B_{i,j}} \frac{1}{h^4} E_{i,j}(z)^2 d\Lambda_2(z)
\]

\[
\leq \int_{B_{i,j}} \left\{ \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right\}^2 d\Lambda_2(z)
\]

\[
+ 2Ch^2 \int_{B_{i,j}} \left| \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right| d\Lambda_2(z) + C^2 h^2
\]

To bound these integrals we use Cauchy-Schwarz:

\[
\int_{B_{i,j}} \left\{ \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right\}^2 d\Lambda_2(z) \leq \int_{B_{i,j}} \left\| \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right\|^2 \left\| \nabla \pi(z) \right\|^2 d\Lambda_2(z)
\]

\[
\leq \frac{h^2}{2} \int_{B_{i,j}} \left\| \nabla \pi(z) \right\|^2 d\Lambda_2(z)
\]

(9)

and

\[
\int_{B_{i,j}} \left| \left( \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right) \cdot \nabla \pi(z) \right| d\Lambda_2(z) \leq \int_{B_{i,j}} \left\| \frac{h}{2} \left[ \begin{array}{c} 2i - 1 \\ 2j - 1 \end{array} \right] - z_2 \right\| \left\| \nabla \pi(z) \right\|_2 d\Lambda_2(z)
\]

\[
\leq \frac{h}{\sqrt{2}} \int_{B_{i,j}} \left\| \nabla \pi(z) \right\|_2 d\Lambda_2(z).
\]

(10)

Both expressions in (9) and (10) are finite since the integrand is continuous and the domain is compact. The total integrated bias is thus bounded as

\[
\int_{Z^2} b^2 d\Lambda_2 \leq \frac{h^2}{2} \int_{Z^2} \left\| \nabla \pi(z) \right\|^2_2 d\Lambda_2(z) + C^2 h^2 + O(h^3)
\]

To bound the variance term, from the integral form of the mean value theorem we have that, for some $z_{i,j} \in B_{i,j}$,

\[
p_{i,j} = \int_{B_{i,j}} \pi d\Lambda_2 = h^2 \pi(z_{i,j}).
\]
The application of the integral form of the mean value theorem is valid since $\pi$ is continuous on $\mathcal{Z}_2$. Then:

$$
\int_{\mathcal{Z}_2} v^2 d\Lambda_2 = \sum_{i,j=1}^{M} \int_{B_{i,j}} v d\Lambda_2
$$

$$
= \sum_{i,j=1}^{M} \int_{B_{i,j}} p_{i,j}(1 - p_{i,j}) \frac{1}{nh^4} d\Lambda_2
$$

$$
= \frac{1}{nh^2} \sum_{i,j=1}^{M} p_{i,j}^2
$$

$$
= \frac{1}{nh^2} \sum_{i,j=1}^{M} \pi(z_{i,j})^2
$$

$$
= \frac{1}{nh^2} - \frac{1}{n} \left( \int_{\mathcal{X}_2} \pi^2 d\Lambda_2 + o(1) \right) = \frac{1}{nh^2} + O\left(\frac{1}{n}\right)
$$

Putting this all together to obtain a bound:

$$
\mathbb{E}\|\pi - \kappa(S^{(n)})\|_{L^2(\Lambda_2)}^2 = \int_{\mathcal{X}_2} b^2 d\Lambda_2 + \int_{\mathcal{Z}_2} v d\Lambda_2 \quad \text{(Fubini)}
$$

$$
\leq \frac{h^2}{2} \int_{\mathcal{Z}_2} \|\nabla \pi(z)\|_2^2 d\Lambda_2(z) + C^2 h^2 + O(h^3) + \frac{1}{nh^2} + O\left(\frac{1}{n}\right) \tag{11}
$$

where $\mathbb{E}$ denotes expectation with respect to sampling of the data $S^{(n)} \sim \Pi$. From inspection of (11), the estimator error vanishes provided that $h$ is chosen such that $nh^2 \to \infty$. Since convergence in expectation implies convergence in probability, we have established that $\|\pi - \kappa(S^{(n)})\|_{L^2(\Lambda_2)} = o_P(1)$. The bandwidth $h^*$, which minimizes the upper bound in (11), is

$$
h^* = \frac{1}{n^{1/4}} \left( \frac{2}{\int_{\mathcal{Z}_2} \|\nabla \pi(z)\|_2^2 d\Lambda_2(z) + 2C^2} \right)^{1/4}
$$

and with this choice we have that $\mathbb{E}\|\pi - \kappa(S^{(n)})\|_{L^2(\Lambda_2)}^2 = O_P(n^{-1/2})$. For $h = h^*$ we have thus established that $\|\pi - \kappa(S^{(n)})\|_{L^2(\Lambda_2)} = O_P(n^{-1/4})$. 

---

**Appendix B. Consistency of the Classifier**

Let $\mathcal{X}$ be the compact metric space $\mathcal{X} = \times_{1\leq i,j \leq M}[0,n]$ from the main text, where $n$ (the number of points in each scatter plot) is fixed. Let $\mathcal{Y} = \mathbb{R}$, so that $\{-1,+1\} \subset \mathcal{Y}$. This section studies the performance of the classifier $\hat{c} : \mathcal{X} \to \{-1,+1\}$, $\hat{c}(x) = \text{sign}(f)$, where $f$ is the Laplacian-regularized least squares method from (4) in the main text, trained on labelled data $\{(x_{[k]}, y_{[k]}): [k] \in \mathcal{L}\}$ and unlabelled data $\{x_{[k]}: [k] \in \mathcal{U}\}$, where $x_{[k]} \in \mathcal{X}$ and $y_{[k]} \in \mathcal{Y}$. To this end, we must establish a context in which the data pairs $(x_{[k]}, y_{[k]})$ can be considered to be generated. Let $\rho_{\mathcal{X},\mathcal{Y}}$ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$, with marginals...
$\rho_X, \rho_Y$ and conditional $\rho_{Y|X}$. In this theoretical investigation we suppose that all data are generated independently from $\rho_{X,Y}$, with the values $\{y[k] : [k] \in \mathcal{U}\}$ being withheld.

For a generic classifier $c : \mathcal{X} \to \{-1, +1\}$, define the misclassification rate

$$\mathcal{R}(c) = \frac{1}{2} \int |y - c(x)| d\rho_{X,Y}(x,y).$$

This is minimized by $c_\rho(x) := \text{sign}(f_\rho(x))$ where $f_\rho : \mathcal{X} \to \mathcal{Y}$ is the (typically unavailable) regression function

$$f_\rho(x) = \int y \, d\rho_{Y|X}(y|x).$$

Thus the quantity $\mathcal{R}(c_\rho)$ captures the intrinsic difficulty of the classification task. A classifier $\hat{c}$ is said to be consistent (either in expectation, with high probability, etc.) if $\mathcal{R}(\hat{c}) \to \mathcal{R}(c_\rho)$ in the limit $m_L \to \infty$ of infinite labelled data (with convergence either in expectation, with high probability, etc.). Our consistency argument is based around the following straightforward bound:

**Lemma 6** Fix $\epsilon > 0$ and let $\mathcal{X}_\epsilon := \{x \in \mathcal{X} : |f_\rho(x)| < \epsilon\}$. Then

$$\mathcal{R}(\hat{c}) \leq \mathcal{R}(c_\rho) + \rho_X(\mathcal{X}_\epsilon) + \frac{1}{2\epsilon} \|f - f_\rho\|_{L^1(\rho_X)},$$

where $\rho_X(\mathcal{X}_\epsilon)$ denotes the $\rho_X$-measure of the set $\mathcal{X}_\epsilon$.

**Proof** For all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, we have that

$$|y - \text{sign}(\hat{f}(x))| \leq |y - \text{sign}(f_\rho(x))| + |\text{sign}(f_\rho(x)) - \text{sign}(\hat{f}(x))|$$

so in particular

$$\mathcal{R}(\hat{c}) \leq \mathcal{R}(c_\rho) + \frac{1}{2} \|\text{sign}(f_\rho) - \text{sign}(\hat{f})\|_{L^1(\rho_X)}, \quad (12)$$

Now,

$$\|\text{sign}(f_\rho) - \text{sign}(\hat{f})\|_{L^1(\rho_X)} = \int_{\mathcal{X}_\epsilon} |\text{sign}(f_\rho) - \text{sign}(\hat{f})| d\rho_X + \int_{\mathcal{X}\setminus\mathcal{X}_\epsilon} |\text{sign}(f_\rho) - \text{sign}(\hat{f})| d\rho_X. \quad (\ast)$$

To bound $(\ast)$, we note that the integrand is trivially bounded by $2$. To bound $(\ast\ast)$, we note that if $|f_\rho(x)| > \epsilon$ then $\text{sign}(f_\rho) \neq \text{sign}(\hat{f})$ implies that $|\hat{f}(x) - f_\rho(x)| > 2\epsilon$. Thus

$$\ast \ast \leq 2\rho_X(\mathcal{X}_\epsilon) + \int_{\mathcal{X}\setminus\mathcal{X}_\epsilon} \frac{|f_\rho(x) - \hat{f}(x)|}{\epsilon} d\rho_X(x) = 2\rho_X(\mathcal{X}_\epsilon) + \frac{1}{\epsilon} \|\hat{f} - f_\rho\|_{L^1(\rho_X)} \quad (13)$$

Combining (12) and (13) completes the proof. 

Next we leverage an existing high-probability consistency result established in the regression (as opposed to classification) context:
Theorem 7. Suppose \( f_\rho \) is non-constant and that \( \Sigma_{K}^{-1/2} f_\rho \in L^2(\rho_X) \) for some \( \alpha \in (0,1] \). Let \( \theta = \frac{1}{(1+\alpha)(1+s)} \). Take \( \lambda_1 = m_U^0 \) and \( \lambda_2 = m_L^0 \). Then there exists a finite constant \( C \) such that for any \( \delta \in (0,1) \), and for \( m_L, m_U \) sufficiently large, we have with probability at least \( 1 - 8\delta \) that
\[
\| \hat{f} - f_\rho \|_{L^1(\rho_X)} \leq C \log \left( \frac{2}{\delta} \right) m_L^{-\alpha\theta}.
\]

Proof. This result is an immediate consequence of Theorem 5.6 in Cao and Chen (2012), whose bound on the \( L^2(\rho_X) \) error clearly also implies a bound on the \( L^1(\rho_X) \) error. In addition, since our intention in what follows is limited to establishing consistency of the proposed classification method, as opposed to a detailed convergence rate analysis, we have simplified the presentation by stating a slightly weaker but less-verbose upper bound.

Note how the “for \( m_U \) sufficiently large” condition in Theorem 7 will typically be automatically satisfied in our context, where the amount of unlabelled data is \( m_U = O(p^2) \). Thus the content of (14) is control over \( \hat{f} - f_\rho \) as the number \( m_L \) of labeled data is increased.

Corollary 8. Under the same assumptions as Theorem 7, we have with probability at least \( 1 - 8\delta \) that
\[
R(\hat{c}) \leq R(c_\rho) + \rho_X(X_\epsilon) + C \frac{2}{2\epsilon} \log \left( \frac{2}{\delta} \right) m_L^{-\alpha\theta}.
\]

Corollary 8 makes explicit how the intrinsic difficulty of the classification task depends on the form of \( f_\rho \), and in particular the extent to which \( |f_\rho(x)| < \epsilon \) occurs in \( X \). For typical regression functions \( f_\rho \) with simple roots in \( X \), it will hold that \( \rho_X(X_\epsilon) = O(\epsilon) \). An assumption of this form can therefore be used to complete a high probability consistency argument:

Corollary 9 (Consistency of the Classifier). Suppose that \( \rho_X(X_\epsilon) = O(\epsilon^\gamma) \) for some \( \gamma > 0 \). Under the same assumptions as Theorem 7, there exists a finite constant \( \tilde{C} \) such that, with probability at least \( 1 - 8\delta \),
\[
R(\hat{c}) \leq R(c_\rho) + \tilde{C} \left( \log \left( \frac{2}{\delta} \right) \right)^{\frac{1}{1+\gamma}} m_L^{-\frac{\alpha\theta}{1+\gamma}}.
\]

In particular, this establishes that the classifier \( \hat{c} \) is (with high probability) consistent.

Proof. From the hypothesis, \( \exists B_1, \epsilon_1 \) such that \( \rho_X(X_\epsilon) \leq B_1 \epsilon^\gamma \) for all \( \epsilon < \epsilon_1 \). Thus, for \( \epsilon < \epsilon_1 \) the difference \( R(\hat{c}) - R(c_\rho) \) can be bounded via (15) as
\[
R(\hat{c}) - R(c_\rho) \leq \rho_X(X_\epsilon) + C \frac{2}{2\epsilon} \log \left( \frac{2}{\delta} \right) m_L^{-\alpha\theta} \\
\leq B_1 \epsilon^\gamma + \frac{B_2}{\epsilon} =: J(\epsilon)
\]
where \( B_2 = \frac{C}{2} \log \left( \frac{2}{\delta} \right) m_L^{-\alpha \theta} \). Differentiating \( J \) and setting to zero reveals that \( J \) is minimized over \((0, \infty)\) at

\[
\epsilon^* = \left( \frac{B_2}{\gamma B_1} \right)^{\frac{1}{1+\gamma}},
\]

which satisfies \( \epsilon^* < \epsilon_1 \) for \( m_L \) sufficiently large (recall that \( m_L \) being sufficiently large was an assumption of Theorem 7). Thus, for \( m_L \) sufficiently large,

\[
\mathcal{R}(\hat{c}) - \mathcal{R}(c_{\rho}) \leq J(\epsilon^*) = \left( \gamma^{-\frac{1}{1+\gamma}} + \gamma^{\frac{1}{1+\gamma}} \right) B_1^{\frac{1}{1+\gamma}} B_2^{\frac{1}{1+\gamma}}
\]

which, upon substitution for \( B_2 \), yields the required result with the value for the constant \( \tilde{C} = \left( \gamma^{-\frac{1}{1+\gamma}} + \gamma^{\frac{1}{1+\gamma}} \right) B_1^{\frac{1}{1+\gamma}} \left( \frac{C}{2} \right)^{\frac{1}{1+\gamma}}. \)
Appendix C. Additional Figures

Figure 6: Data set D1 (yeast data), random sampling. ROC curves are shown with respect to causal relationships determined from unseen interventional data (see Main Text for details). Results for PC (which returns a point estimate) are shown as locations on the ROC plane. “TC” indicates use of a transitive closure operation and “cnstrnts” indicates that the background information Φ was included via input constraints. [Results shown are for significance level α = 0.01 and for a lenient interpretation where possible edges are included. Results are averages over 25 iterations.]
Figure 7: Data set D1 (yeast data), row-wise sampling. ROC curves are shown with respect to causal relationships determined from unseen interventional data (see Main Text for details). Results for PC (which returns a point estimate) are shown as locations on the ROC plane. “TC” indicates use of a transitive closure operation and “cnstrnts” indicates that the background information $\Phi$ was included via input constraints. [Results shown are for significance level $\alpha = 0.01$ and for a lenient interpretation where possible edges are included. Results are averages over 25 iterations.]
Figure 8: Data set D3 (cancer protein data). ROC curves are shown with respect to a reference graph determined from the scientific literature (see Main Text for details). Results for PC (which returns a point estimate) are shown as locations on the ROC plane. “TC” indicates use of a transitive closure operation and “cnstrnts” indicates that the background information $\Phi$ was included via input constraints. The “TC” results are included here for completeness, but we note that the reference graph here encodes direct, rather than ancestral, relationships. [Results shown are for significance level $\alpha = 0.01$ and for a lenient interpretation where possible edges are included. Results are averages over 25 iterations.]